



**US Army Corps
of Engineers®**
Engineer Research and
Development Center



Socio-economic Effects of Drought in the Horn of Africa

Population Movements, Livelihoods, Market Prices, and Infrastructure

Jeanne M. Roningen and John B. Eylander

April 2014

The US Army Engineer Research and Development Center (ERDC) solves the nation's toughest engineering and environmental challenges. ERDC develops innovative solutions in civil and military engineering, geospatial sciences, water resources, and environmental sciences for the Army, the Department of Defense, civilian agencies, and our nation's public good. Find out more at www.erdclibrary.usace.army.mil.

To search for other technical reports published by ERDC, visit the ERDC online library at <http://acwc.sdp.sirsi.net/client/default>.

Socio-economic Effects of Drought in the Horn of Africa

Population Movements, Livelihoods, Market Prices, and Infrastructure

Jeanne M. Roningen and John B. Eylander

Cold Regions Research and Engineering Laboratory (CRREL)
US Army Engineer Research and Development Center
72 Lyme Road
Hanover, NH 03755-1290

Final Report

Approved for public release; distribution is unlimited.

Prepared for The National Geospatial-Intelligence Agency
7500 GEOINT Drive
Springfield, Virginia 22150-7500

Under National Geospatial-Intelligence Agency Military Interdepartmental Purchase
Request No: NIB8G12206GSA9

Abstract

In recent years, the intelligence and defense communities have indicated interest in understanding the potential relationships between anomalous climate events and socio-economic consequences outside of the United States that could have implications for US national security. Our research evaluates potential linkages between retrospective climate analyses and empirical socio-economic datasets in Somalia and Kenya surrounding the 2011 drought in the Horn of Africa. Subnational-level data on internally displaced persons in Somalia from 2008–2012 were used to correlate drought-related population movements to climate-model-derived moisture indices. The analysis was expanded to account for livelihood zones and to investigate the predictive capabilities of linear models for observed population movements. Additional analyses investigated market price response to drought and market connectivity and explored the use of census data on household water infrastructure to assess drought vulnerability of specific communities. Results suggest that drought-induced migration response occurred with low but significant correlations across a broad range of medium- to long-term (6 months to 4 years) standardized drought indices but was limited largely to the geographic area in southern Somalia subject to a confluence of three factors: al-Shabaab governance during the 2011 drought, associated legal and operational impediments to aid delivery, and non-arid seasonality patterns.

Contents

Abstract	iv
Illustrations	vii
Preface	xi
Nomenclature	xii
Executive Summary	xiv
1 Introduction	1
1.1 Climate disruptions and US security interests	1
1.2 Study objective	2
1.3 Study area background	3
1.3.1 Drought in Somalia, 2010–2011	3
1.3.2 Drought in Kenya, 2010–2011	6
2 Methods	8
2.1 Overview	8
2.2 Climate data	8
2.2.1 Standardized precipitation index	10
2.2.2 Standardized soil water index	12
2.2.3 Benefits of standardized indices	13
2.2.4 Additional considerations when using standardized indices	13
2.3 Drought and population movement	14
2.3.1 Population movement tracking data	14
2.3.2 Spatial resolution and alignment of gridded and polygon datasets	15
2.3.3 Correlation methods	17
2.3.4 Livelihood zones	20
2.3.5 Linear models	21
2.4 Drought and market prices	23
2.4.1 Market price data	23
2.4.2 Market price ratio to moisture index correlations	25
2.4.3 Cross-correlation functions	25
2.4.4 Network analysis	27
2.5 Census data and vulnerability to drought	28
2.5.1 Mapping of census data	30
2.5.2 Association rules analysis	30
3 Results and Discussion	33
3.1 Overview	33
3.2 SPI and SSWI indices	33
3.3 Correlations between drought-induced IDPs and moisture indices	36

3.3.1 Overview and discussion of PMT data	36
3.3.2 Initial data analysis.....	39
3.3.3 Correlations between IDPs and moisture indices using Pearson's r	41
3.3.4 Correlations between IDPs and moisture indices using Kendall's τ	45
3.3.5 Correlations between PMT and moisture indices by livelihood zone	48
3.3.6 Linear models.....	52
3.4 Market price network analysis.....	62
3.4.1 Market price trends	63
3.4.2 Correlations between prices and moisture indices.....	66
3.4.3 Lag times for market prices and drought.....	69
3.4.4 Inter-market price correlations.....	74
3.4.5 Degree centrality.....	79
3.5 Census data	81
3.5.1 Mapping of census data.....	81
3.5.2 Drought locations in Kenya	85
3.5.3 Association rules analysis	86
4 Conclusions and Recommendations	91
4.1 MERRA-Land.....	92
4.2 SPI/SSWI indices.....	92
4.3 Drought and PMT	93
4.4 Drought and FSNAU market prices	95
4.5 Drought and Kenyan census data	96
4.6 R and open-source libraries	96
References	98
Report Documentation Page	

Illustrations

Figures

1	Map of Somalia.....	4
2	Map of Kenya	7
3	Four SPI time series from one grid cell location over the Horn of Africa. Red indicates below-average precipitation at that location over the period of record	11
4	Three-month SPI values from December 2010 to November 2011, based on MERRA-Land precipitation data over Somalia from 1980 to 2012, with administrative districts outlined in black. Note that according to the MERRA-Land dataset, severe drought conditions were not limited to the southern part of the country and that the 2011 late-season minor rains (<i>Deyr</i>) were above normal in the south, contributing to the end of the drought.....	12
5	MERRA-Land grid cells are displayed together with Somali administrative districts (black lines). Using the majority area method, each Somali administrative district was associated with between one and seven MERRA grid cells, and some grid cells were associated with multiple districts	16
6	Livelihood zones in Somalia with MERRA-Land grid cells in the background.....	21
7	FEWSnet seasonal calendar for Somalia.....	22
8	The 1-month Standardized Precipitation Index changes rapidly with changing precipitation values.....	34
9	The 8-month Standardized Soil Water Index quantifies longer-term seasonal anomalies in modeled root zone soil moisture	35
10.	Numbers of drought-induced internally displaced people, by district of out-migration	37
11	Numbers of insecurity-induced internally displaced persons, by district of out-migration. Note that the scale in this figure is approximately one order of magnitude greater than the scale in Figure 10.....	38
12	Comparison for one grid cell in Somalia of 1-month SPI and drought-induced IDPs with good correlation	39
13	Comparison for one grid cell in Somalia of 1-month SPI and drought-induced IDPs exhibiting an apparent lag in drought response of several months.....	40
14	Comparison for one grid cell in Somalia of 1-month SPI and drought-induced IDPs exhibiting apparent IDP response but minimal to no drought	40
15	Comparison for one grid cell in Somalia of 1-month SPI and drought-induced IDPs with apparent anticipatory movement	41
16	Box-and-whisker diagram showing values of Pearson's r between moisture indices and drought-induced population movements across Somalia between 2008 and 2012. Black horizontal lines indicate median correlation values at each moisture index; boxes indicate 1 st and 3 rd quartiles; whiskers are drawn at 1.5 times the interquartile range; circles indicate correlation values that fall outside of this range. The hourglass indentations signify the distance of separation required for two median values to be considered significantly different at a 95% confidence level	42

17	Map of Pearson correlation values between drought-induced IDPs and moisture indices.....	43
18	Significance of mapped Pearson correlation values	44
19	A histogram and quantile-quantile normal plot of IDP movement data in one example district. If normally distributed, the data in the quantile-quantile normal plot would follow a diagonal line through the center of the bottom graph. The graphs show that IDP data is heavily skewed with multiple tied values of zero	45
20	Box-and-whisker diagram showing values of Kendall's τ between moisture indices and drought-induced population movements across Somalia between 2008 and 2012. Median correlation values at each moisture index are indicated with horizontal black line; 1 st and 3 rd quartiles are indicated by boxes; whiskers are drawn at 1.5 times the interquartile range; circles indicates indicate correlation values that fall outside of this range. The hourglass indentations signify the distance of separation required for two median values to be considered significantly different at a 95% confidence level.....	46
21	Map of Kendall correlation values between drought-induced IDPs and moisture indices.....	47
22	Significance of mapped Kendall correlation values	48
23	Correlations between population movements and the 8-month SPI, by livelihood zone	49
24	Maps of mean correlations across livelihood zones in southern Somalia	50
25	Aggregated livelihood zones by major types.....	51
26	The north–south distinction between correlation strengths does not appear to be related directly to aggregated livelihood zone types depicted in Figure 23.....	52
27	Percent of district population that became IDPs due to drought between December 2010 and November 2011 and corresponding livelihood zone-weighted moisture indices	53
28	Open-source map of areas of al-Shabaab control in Somalia, December 2009 (left) and January 2010 (right) (Kermanshahi 2010).....	54
29	Two linear models for drought-induced IDPs as a function of a weighted moisture index distinguished by governance regime.....	55
30	Seasonality map of Africa classified by number and modality of wet seasons and derived from remotely-sensed rainfall and temperature data (Herrmann and Mohr, 2011)	58
31	The 2-month SPI weighted model was run in predictive mode to depict confidence intervals.....	59
32	Four methods for assessing outliers in a linear model	61
33	Residuals map for the livelihood-zone-weighted linear model of drought-induced population response to mean 1-month SPI in al-Shabaab-controlled areas. The scale bar depicts residuals as fraction of total district population. The districts with the highest and lowest residuals were, respectively, a district with a low population where percent population displacement was underestimated by the model (dark green) and a district on the Ethiopian border that was not entirely under al-Shabaab control and where population displacement was overestimated by the model (dark pink)	62
34	Prices for cereals and legumes, 2010–2011 Monetary denominations vary.....	64
35	Market prices for livestock, 2010–2011. Monetary denominations vary	65

36	Market prices for fuel and the day labor rate and the exchange rate for the US dollar and the Somali Shilling, 2010–2011. Monetary denominations vary.....	66
37	Southern Somalia in particular exhibits strong negative correlations indicating price-ratio increases with drought between December 2010 and November 2011.....	68
38	A CCACF analysis appears to demonstrate short-term effects of drought on cattle prices in southern Somalia during the 2011 drought year. However, this initial analysis may not be statistically robust due to edge effects of the CCACF analysis described below.....	70
39	The cross-correlation function for local-quality cattle in one particular market peaks concurrently (lag = 0) with the 2-month SSWI index at that market location	70
40	For the period covering the 2011 drought, the largest deviations from the mean occur near the ends of the time series. This has negative implications for the robustness of the CCACF analysis.....	71
41	Raw data for the 2-month SSWI and FSNAU price-to-day-labor rate for local-quality cattle.....	72
42	Price ratio response to the 2-month SSWI for local-quality cattle from July 2002 through August 2012. In markets closer to Mogadishu, price ratios increased during dry periods; but in markets further from the port city, price ratios decreased.....	73
43	Price ratio response to 2-month SSWI for export-quality cattle from July 2002 through August 2012. The pattern of spatially-differentiated price ratios during drought is similar to that of local-quality cattle, but the variance in lag times is larger	73
44	Market locations in Somalia are mapped as nodes on a network graph. Each graph edge connects two nodes that exhibit a statistically significant Pearson's r correlation of a minimum of 0.5, with darker lines indicating stronger correlations. The two nodes on the left-most side of these and subsequent graphs represent the two market locations whose geographic coordinates were unknown	75
45	Correlation networks for white maize and cowpeas, December 2008–November 2009	75
46	Correlation networks for white maize and cowpeas, December 2009–November 2010.....	76
47	Correlation networks for white maize and cowpeas, December 2010–November 2011.....	76
48	Production and market flow map for maize (FEWSnet 2008)	77
49	Correlation networks for export quality sheep and local quality camels, December 2008–November 2011.....	78
50	Normalized degree centrality for thirteen market products, December 2010–November 2011	79
51	Normalized degree centrality summed across all market products, December 2010–November 2011	80
52	Most frequent attributes of Kenyan census data from 1999	81
53	Rural household fraction	82
54	Household ownership fraction.....	83
55	Household ownership categories.....	83

56	Household rental status	84
57	Household water infrastructure.....	84
58	Household roofing material.....	85
59	Three-month SPI across Kenya, January 2010–August 2012.....	86

Tables

1	Market products monitored by FSNAU (2013)	24
2	Select Kenyan household census variables	29
3	Summary of significance of slope values for different linear models. Green and yellow cells indicate which indices are associated with a slope significantly different from zero (to the 95% and 90% level, respectively). White cells indicate that there is no significant relationship in this analysis for that moisture index. The linear models could be used to estimate potential drought-induced IDPs as a function of a weighted moisture index	56
4	Association rules for rural households in the Marsabit district.....	88
5	Association rules for urban households in the Marsabit district	89
6	Association rules for urban Marsabit households that rent from the government	89
7	Association rules for urban Marsabit households that rent from the government and have asbestos roofs	89

Preface

This study was conducted for the National Geospatial-Intelligence Agency (NGA) under Military Interdepartmental Purchase Request Number NIB8G12206GSA9, “A New Seasonal Drought Prediction Capability for Food Security and Regional Stability Operations.” The technical monitor was Dr. Christy Crosiar, NGA Innovision.

The work was performed by Jeanne M. Roningen (Remote Sensing/GIS and Water Resources Branch, Timothy Pangburn, Chief) and John B. Eylander (Terrestrial and Cryospheric Sciences Branch, Janet Hardy, Chief), US Army Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory (ERDC-CRREL). At the time of publication, Dr. Justin Berman was Chief of the Research and Engineering Division. The Deputy Director of ERDC-CRREL was Dr. Lance Hansen, and the Director was Dr. Robert Davis.

COL Jeffrey R. Eckstein was the Commander of ERDC, and Dr. Jeffery P. Holland was the Director.

Nomenclature

CCACF	Cross-Correlation Function
CPC	Climate Prediction Center
CRREL	US Army Cold Regions Research and Engineering Laboratory
ERDC	Engineer Research and Development Center
FAO	Food and Agriculture Organization
FEWSnet	Famine Early Warning System Network
FSNAU	Food Security and Nutrition Analysis Unit
GEOS-5	Goddard Earth Observing System 5
GES DISC	Goddard Earth Sciences Data and Information Services Center
GMAO	Global Modeling and Assimilation Office
IDP	Internally Displaced Person
IPUMS	Integrated Public Use Microdata Series
KHS	Kenyan Shilling
LNGO	Local Non-Governmental Organization
MERRA	Modern-Era Retrospective Analysis for Research and Applications
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NDVI	Normalized Differential Vegetation Index
NGA	National Geospatial-Intelligence Agency
NGO	Non-Governmental Organization
NOAA	National Oceanic and Atmospheric Association

ORNL	Oak Ridge National Laboratory
PMT	Population Movement Trends
SPI	Standardized Precipitation Index
SSWI	Standardized Soil Wetness Index
TRMM	Tropical Rainfall Measuring Mission
UN	United Nations
UN-OCHA	UN Office for Coordination of Humanitarian Affairs
UNHCR	United Nations High Commissioner for Refugees
USAID	US Agency for International Development
USD	US Dollar
WFP	World Food Programme

Executive Summary

In recent years, the intelligence and defense communities have indicated interest in understanding the potential relationships between anomalous climate events and socio-economic consequences outside of the United States that could have implications for US national security. In 2013, the National Academy of Sciences recommended supporting research “to improve understanding of the conditions under which climate-related natural disasters and disruptions of critical systems of life support do or do not lead to important security-relevant outcomes such as political instability, violent conflict, humanitarian disasters, and disruptive migration”. Our research aimed to use empirical methods to quantitatively evaluate potential linkages between retrospective climate analyses and existing social and economic datasets from the Horn of Africa surrounding the 2011 drought.

Possible connection between drought and instability in Somalia

Data on internally displaced persons (IDPs) in southern Somalia suggest that there may have been a significant, if indirect, causal connection between climate extremes and insecurity following the 2011 drought. The population movement data support observations made by others that the reigning governing body (al-Shabaab) suffered a loss of popular support because of its inability to provide for people during the drought. Aid access was limited by different but related mechanisms: by operational challenges to distribution presented by the security situation; by the local regime itself, in an attempt to foment independence from international aid; and also by anti-terrorism legislation preventing aid delivery via specified organizations, of which the regime was one. The regime’s weak response to the drought is thought to be one of several factors that led to al-Shabaab’s ousting from large parts of southern Somalia in the following year. The figure below shows the locations and dates of drought-caused and insecurity-caused IDPs, with an arrow representing the possible causal connection. These data are preliminary to the current study, rather than a result of it, but address the fundamental question of how a historical climate event might have impacted a conflict situation in the study area. Although the numbers of insecurity-induced movements are an order of magnitude higher than the drought-induced movements, if there

is a causal connection then the drought-induced movements may have an outsized importance.



Broad swaths of IDPs resulted from the failed rainy seasons of the 2011 drought. These IDPs precede and may have had a causal connection to insecurity-related IDPs associated with the ousting of al-Shabaab the following year in large parts of southern Somalia.

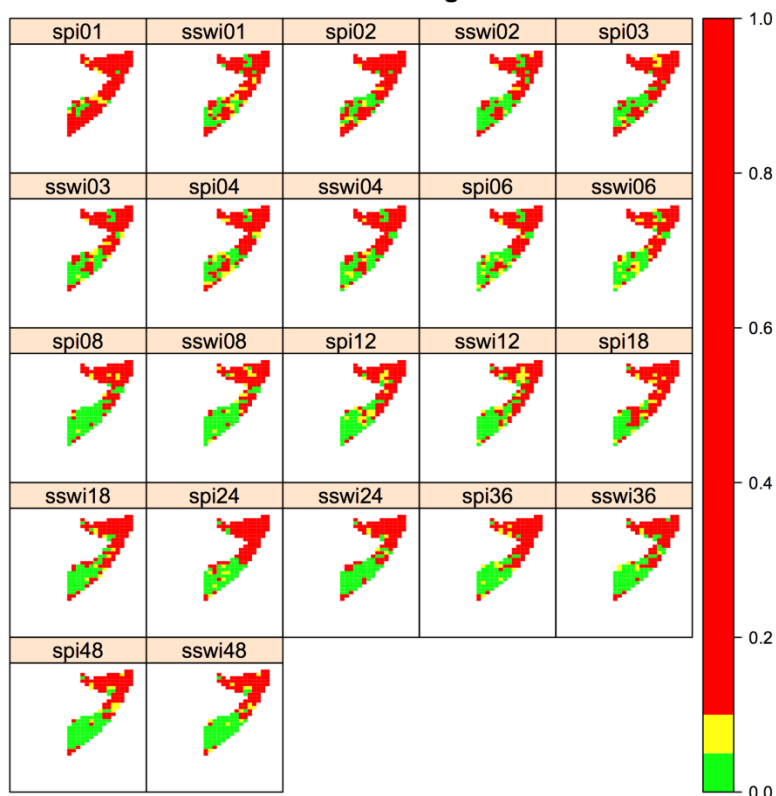
Any such climate–security link invites open discussion surrounding the possibility that climate events could therefore be used strategically by multiple actors. There may be an inherent conflict of interest between humanitarian actors wishing to respond to affected persons in the short term and military and political actors who, to shape public opinion and influence military or political outcomes, may wish to limit certain governing groups' capability to respond to climate-related stress.

Spatial and temporal thresholds for population movement response to drought in Somalia

In southern Somalia, standardized indices that measured the drought intensity over the previous 6 to 48 months had statistically significant correlations to population movements. This showed the relationship between drought and drought-induced population movements in this region over the medium to long term, as indicated by the green areas in the figure below. However, the magnitudes of the correlation values were generally weak, confirming the well-known fact that drought itself, or at least the drought represented in the climate dataset used, is not the sole

driver of drought-induced population movement in these areas. Interestingly, there was a notable lack of significant correlation in the corridor surrounding the capital and port, Mogadishu, for moisture anomalies calculated over the 2–6 month time frames, potentially indicating the expectations of commerce or aid from the port city as the drought developed.

P-values for Correlations between Drought IDPs and Moisture Indices



Grid cells in green indicate where Kendall's tau correlations are significant at the 95% level (two-sided)

Green areas show evidence of statistically significant correlations between the moisture index and drought-induced population movements.

In the northern half of the country, drought indices were only sporadically correlated with population movements. This region differs from southern Somalia in at least two significant respects: its climate is routinely arid, so moisture anomalies may be less disruptive to the economy; and it was not governed by al-Shabaab and, therefore, was not subject to restrictions on aid associated with insecurity or anti-terrorism legislation. To account for both regular dry (or wet) seasons and year-round arid (or rainy) climates, it is possible that some combination of anomaly-based and absolute moisture indices will be better suited for this type of analysis than anomaly-based indices, such as those used in this study, alone.

Additional results

In addition to the above, this study explored multiple techniques in conjunction with livelihood, market price, and census data sets in the Horn of Africa surrounding the 2011 drought.

- We extended the population movement analysis to incorporate seasonal calendar information based on livelihood zones into linear models of population movement during the drought. The weighted models showed an improved correspondence in population movement response to short-term (2–4 month) precipitation-based moisture indices as compared to similar unweighted linear models. Models using short-term indices based on soil moisture did not show any significant correspondence with or without weighting.
- Some market-price-analysis techniques showed patterns in staple grain and livestock that appeared to correspond to geographic patterns in population movement data and to proximity to the port of Mogadishu. It was not clear from this research alone to what extent these market price analyses may offer opportunities to improve upon existing expert or local understanding of the social, economic, or security situation brought about by drought in this region. Additional input from humanitarian or security communities would be required to assess whether local variability in these datasets may be indicative of some specific security or aid situation on the ground or of external economic factors.
- The technique of association rules analysis was successfully applied to Kenyan census data to characterize subcommunities in drought-affected areas by their access to household water infrastructure. However, this data was not sufficient alone to draw conclusions about community vulnerability to drought without further ground-based knowledge as it did not include information on the actual functioning of the infrastructure.

Our research aims to use empirical data to help understand and visualize evolving complex environmental situations of international security interest. Having developed an infrastructure for this type of analysis, the methods found to be most successful here could also be extended to other climate datasets and phenomena (e.g., for floods rather than for droughts or for normalized differential vegetation indices rather than for

retrospective climate models) and to other spatial and temporal scales of interest with a primary limiting factor being the availability of subnational-scale, socio-economic data. In particular, the climate-population movement correlations used here could help differentiate climate-induced population movements from larger undifferentiated datasets that include population movement from all causes.

1 Introduction

1.1 Climate disruptions and US security interests

In recent years, the intelligence and defense communities have indicated interest in understanding the potential relationships between anomalous climate events and socio-economic consequences outside of the United States that could have implications for US national security (Department of State and USAID 2010; Office of the Director of National Intelligence 2012; Army Research Office 2011). In 2013, the National Academy of Sciences recommended supporting research “to improve understanding of the conditions under which climate-related natural disasters and disruptions of critical systems of life support do or do not lead to important security-relevant outcomes such as political instability, violent conflict, humanitarian disasters, and disruptive migration” (National Research Council 2013).

According to intelligence estimates, some 12 to 15 countries fall into the categories of both having significant potential for climate change impacts and being of importance to US national security in the next ten years; many of these are located in South Asia, the Middle East, and Africa. If the set of countries is expanded to include those for which the US has foreign policy and humanitarian concerns as well as security concerns, the length of the list grows to 50 or 60 countries.

Developing an adequate system for monitoring the conditions that can link climate events to national security concerns will require . . . the analysis of new and existing data and the improvement of analytic systems, leading to better understanding of the linkages over time and to improved indicators of key variables where quantitative indicators are appropriate and feasible to produce. It will typically require finer-grained data than are currently available (National Research Council 2013).

Climate projections alone are not sufficient to understand the ways in which societies and governments will respond to climate events. Rather, responses are determined by a combination of event severity; the exposure

of the population and critical support systems to an event or event sequence; and mechanisms for coping, response, and recovery on the part of the population and its support systems. Many studies of environmental refugees, as an example, are largely based only on population estimates and projected areas of exposure rather than on empirical studies of actual population movements during past or ongoing events (Gemenne 2011). Therefore, “security analysis needs to develop more nuanced understanding of the conditions—largely, social, political, and economic conditions—under which particular climate events are and are not likely to lead to particular kinds of social and political stresses and under which such events and responses to them are and are not likely to lead to significant security threats” (National Research Council 2013).

1.2 Study objective

Our study aimed to use existing datasets on climate and socio-economic phenomena in the Horn of Africa to improve understanding of the conditions under which the drought of 2011 was or was not associated with socio-economic outcomes of potential interest to US security. We separated this objective into six semi-autonomous components:

1. Selecting retrospective climate datasets for historical analysis
2. Selecting appropriate drought indices based on these datasets
3. Assessing quantitative spatial relationships between drought and population movements
4. Assessing quantitative spatial relationships between drought and market prices
5. Assessing the suitability of census data for understanding population vulnerability to drought
6. Developing a set of flexible and extensible techniques for data ingestion and processing for potential use in future analyses.

We selected the study locations of Somalia and Kenya for multiple reasons. First, both countries had regions affected by the Horn-of-Africa-wide drought of 2011. Second, each country had in the public domain one or more datasets on population movements, market prices, or population characteristics that were suitable for statistical analysis at subnational spatial resolutions. Third, a simultaneous climate study is being conducted by the National Aeronautical and Space Agency (NASA) in the Horn of Africa as part of the same program under which our study falls, “A New Seasonal Drought Prediction Capability for Food Security and Regional Stability

Operations.”* The NASA portion of the project aims to improve drought prediction by assimilating satellite-derived soil moisture and other remotely-sensed datasets into a climate model. Our study provides a baseline for further comparisons between the utility of different climate datasets in evaluating the social and economic consequences of drought.

1.3 Study area background

1.3.1 Drought in Somalia, 2010–2011

The Federal Republic of Somalia (Fig. 1) is located in the East African region known as the Horn of Africa and is bordered to the east by the Indian Ocean; to the north by the Gulf of Aden; and inland by neighboring countries Djibouti, Ethiopia, and Kenya. Arid regions in most of the north and a dual wet-season pattern with major (*Gu*) and minor (*Deyr*) rains in the south and parts of the northwest characterize the climate (Herrmann and Mohr 2011). The country has been without a unified federal government since the ousting of the Siad Barre regime in 1991 and has experienced multiple periods of significant instability (Kovac 2013).

On 23 August 2010, the US Agency for International Development’s (USAID) Famine Early Warning System Network (FEWSnet) issued a Somali Food Security Alert report noting that seasonal forecasts using output from multiple climate analyses predicted a La Niña event, which is often related to lower than average seasonal rains in the Horn of Africa between October and December. The report also noted that in four of the last six La Niña events, the major rainy season between March and May in the region was also poor. A month later, in September 2010, another FEWSnet brief put a probability of 50% to 60% on the prediction that March–May in the eastern part of the Greater Horn of Africa would suffer from poor rains.

* The NASA study is being conducted under National Geospatial-Intelligence Agency Military Interdepartmental Purchase Request Number NIB8G12103GS52.

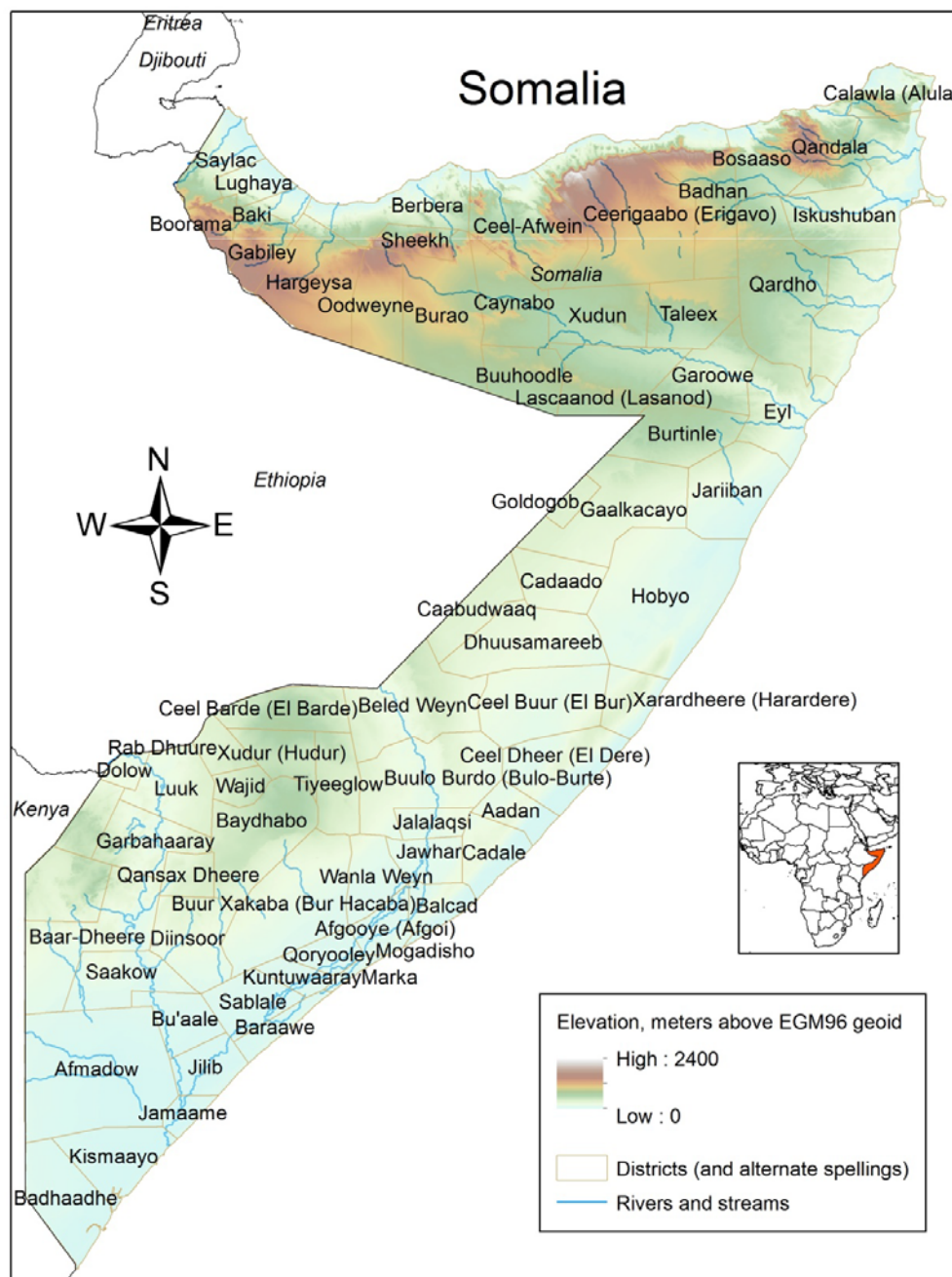


Figure 1. Map of Somalia.

On 20 July 2011, eleven months after the earliest warnings that a food security situation might arise in the Horn of Africa, famine was officially declared in parts of southern Somalia, based on indicators from a specialized biannual assessment of nutrition and mortality by FEWSnet and the United Nation's (UN) Food Security and Nutrition Analysis Unit (FSNAU) (Maxwell et al. 2012). During the peak of the famine, which officially ended in early 2012, four in ten children in southern Somalia were acutely

malnourished, four million people lacked access to basic necessities, and as many as 2000 people per day arrived into refugee camps outside of Somalia's borders (Hillbruner and Moloney 2012). A retrospective analysis of mortality caused by the famine estimates that 258,000 people died of severe malnutrition, half of whom were children under the age of five (Checchi and Robinson 2013).

While drought is often a major proximate or immediate cause of famine, the humanitarian community widely acknowledges famine to be caused in large part by anthropogenic factors rather than solely environmental conditions; poor governance, often associated with conflict, plays a central role. In the Horn of Africa drought of 2011, three proximate factors to the Somali famine were drought, conflict, and food prices. Underlying and complicating factors included longer-term, conflict-induced changes in land access; movement restrictions and forced refugee returns by the militant Islamist group al-Shabaab, which had gained power over large areas of southern Somalia; and US counter-terrorism policies restricting material support and hence aid distribution via foreign terrorist organizations (Maxwell and Fitzpatrick 2012).

Population movements can be both an appropriate response to and an undesirable effect of climate events. Many Somalis derive their livelihoods from pastoral livestock; and as such, both nomadic and semi-nomadic (transhumant) population movements occur seasonally as a normal response to weather conditions. During anomalous climate events, population movement on the part of those most affected is often a useful and appropriate response (National Research Council 2013). However, population movements can also be destabilizing and undesirable when the destination areas are not able to absorb and integrate arrivals. Movements can also exacerbate vulnerability to later climate events as displacements often require the sale of productive assets or the use of financial reserves; reduce ongoing maintenance and investment in livelihood mechanisms, such as riverine water control infrastructure (Little 2008); and sever ties to support networks and social services, such as health, sanitation, and education (Maxwell and Fitzpatrick 2012). Population movements can then have long-term as well as short-term consequences: certain subsets of the Somali population, notably two minority groups within Somalia who had been displaced from their land during the civil war in the early 1990s and who were marginal to or outside of the major clan structures found in the region, were found to have been particularly vulnerable during the

2011 famine, both among rural populations and in Internally Displaced Person (IDP) camps (Majid and McDowell 2012). Following the US-backed Ethiopian invasion of 2006 and prior to the beginning of the drought in late 2010, 1.46 million Somalis had been displaced from their homes largely due to conflict (UNHCR 2012), representing the highest level of civilian displacements since the Somali civil war in the early 1990s (Maxwell and Fitzpatrick 2012). While much of the insecurity-induced displacement was from urban areas, drought-induced displacement occurred largely from agropastoral and pastoral zones (Robinson et al. 2014).

1.3.2 Drought in Kenya, 2010–2011

The neighboring Republic of Kenya (Fig. 2), with its northeastern sections in the Horn of Africa and its western parts in the African Great Lakes region, shares coastline with the Indian Ocean and an eastern border with southern Somalia. It is bordered to the north by Ethiopia and South Sudan, to the west by Uganda, and to the south by Tanzania. Its climate is arid in the northwest Turkana area and in the eastern and less-populated two-thirds of the country experiences a dual wet-season pattern with major and minor rains. The climate in the more populous western highlands of the Great Lakes region is humid (Herrmann and Mohr 2011). The country has been a sovereign entity since 1963, enduring epochs of one-party rule and moderate democratic liberalization without state failure or party overthrow (Hornsby 2013).

In Kenya, the Horn of Africa drought resulted in large losses of livestock in pastoral areas, overstretched refugee camps hosting cross-border refugees (UN-OCHA 2011), and insecurity issues related to smuggling of small arms and weapons concurrent with uncontrolled refugee flows. Unlike Somalia, where years of conflict and absence of a stable national government prevented national-level response, Kenya had basic governmental mechanisms in place to declare and respond to the drought as a national disaster. Funds were used for food provision, refugee camp sanitation, water delivery services, school feeding programs, livestock buy-out programs, and cash transfers to people living in drought-affected areas. However, the government response was subject to criticism for being slow, ineffective, and susceptible to corruption (Opiyo 2011; Tatalovic 2011; Ross 2011). Of primary interest in this study is the feasibility and utility of using existing public census records on household water infrastructure characteristics to identify vulnerable subpopulations in drought-affected areas.

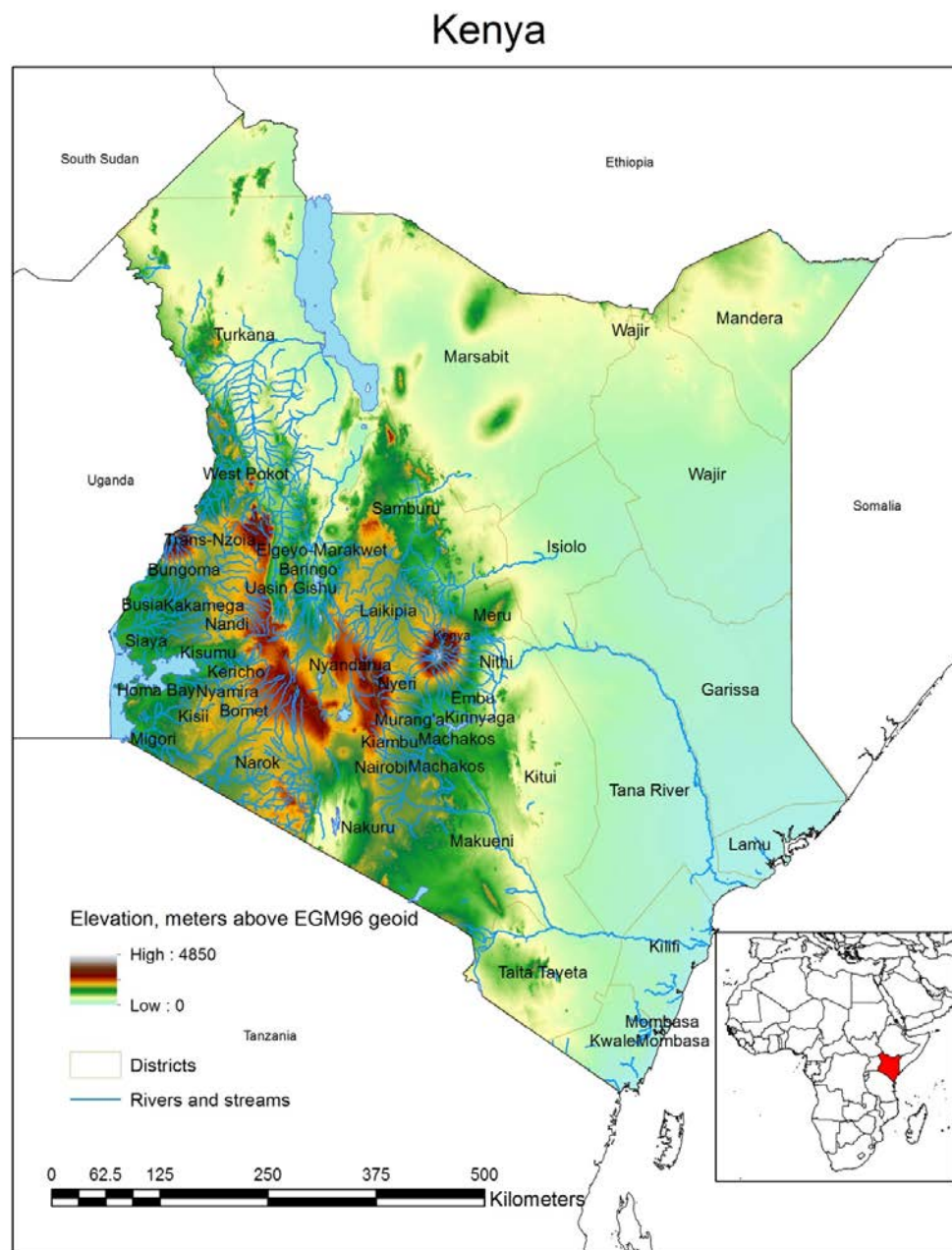


Figure 2. Map of Kenya.

2 Methods

2.1 Overview

Using the methods described below, we addressed the set of six semi-autonomous tasks that compose the study objective. Section 2.2 briefly presents the climate dataset used for historical analysis and outlines the reasons for that choice. Section 2.3 describes the moisture indices used to quantify drought and the reasons for their use. Section 2.4 describes the data sets and methods used to assess potential relationships between these moisture indices and population movements. Section 2.5 describes the data sets and methods used to assess connections between market prices and drought, and Section 2.6 describes the use of census data sets to assess potential vulnerability to drought. We implemented the last task, developing an open-source code base suitable for follow-on analyses, using *R*, an open-source language and environment for statistical computing (R Core Team 2012b). This task does not have a section devoted to it but is described throughout at relevant steps.

2.2 Climate data

The Modern-Era Retrospective Analysis for Research and Applications (MERRA) (Rienecker et al. 2011) is a climate reanalysis produced by NASA's Global Modeling and Assimilation Office (GMAO). A climate reanalysis is a reconstruction of past weather by combining physical weather models (in this case, the Goddard Earth Observing System 5 model [GEOS-5]) and satellite, weather station, and weather balloon observations. The reanalysis produces a reconstruction of past weather at finer spatial and temporal resolutions than that afforded by weather observations alone and uses those observations to constrain the model physics so that the model, which is subject to the same difficulties as any weather forecast system, does not stray too far from observations. The objective of the MERRA dataset in particular was to ensure consistency in the way observations were incorporated into the physical models, even as new observations became available over the years via NASA's Earth Observing System satellites. This effort produced a consistent climate product spanning the period since satellite observations have become available.

In our study, we use a variant of the MERRA dataset: MERRA-Land. MERRA-Land was developed to mitigate some of the known hydrologic biases in the original MERRA data and differs from it in three principal ways. First, the MERRA-Land product was derived from an offline run of the land surface model, rather than a fully coupled version where the land surface model is allowed to provide feedback to the atmospheric model. Second, in this offline run, the atmospheric forcings from the GEOS-5 model included a modified set of precipitation data. This modified precipitation data consisted of original MERRA precipitation that was merged with separate gauge-based precipitation records from the National Oceanic and Atmospheric Association's (NOAA) Climate Prediction Center (CPC). Third, the original MERRA data were calculated using a different version (the "MERRA" version) of the Catchment hydrology land surface model (Koster et al. 2000) whereas MERRA-Land used an updated "Fortuna 2.5" version of the Catchment land surface model. With these modifications, researchers found that MERRA-Land mitigated some of the known hydrologic biases in the original MERRA data (Reichle et al. 2011).

The reasons for our choice of the MERRA-derived datasets from among many available climate and land surface products were that (1) reliable, continuous gauge-based observations for precipitation and soil moisture are extremely sparse in the Horn of Africa; (2) the period of record of the climate data needed to be long enough to provide adequate historical perspective; (3) methods for data assimilation had to be consistently applied over time despite the introduction of new satellite observation systems and assimilation techniques; and (4) a land surface model that included reservoirs for soil and ground water was desired to allow for the possibility of assimilating newer satellite products, such as AMSR-E and GRACE, in future land surface analyses. MERRA products, and the associated MERRA-Land products, simultaneously address all of these concerns. However, it is important to note that we did not conduct any validation of this dataset over the Horn of Africa as a part of this study.

From MERRA-Land, and for reasons explained in the following sections, we used the Total Surface Precipitation and Root Zone Soil Wetness products, two of multiple MERRA-Land "Monthly IAU 2d Simulated land surface diagnostics" outputs. These products were subset over the geographic area of interest at monthly intervals and downloaded as netcdf files from the Goddard Earth Sciences Data and Information Services Center (GES DISC 2010). The grid cell size of the dataset is $1/2^\circ$ latitude by $2/3^\circ$ longi-

tude, translating to cell sizes of approximately 55 km north to south by 75 km east to west near the equator.

2.2.1 Standardized precipitation index

All moisture indices used in this study are based on the same method used to calculate the Standardized Precipitation Index (SPI) (McKee et al. 1993). The SPI sums precipitation over a given duration (for example, August and the preceding two months in a given year) and compares it to other cumulative precipitation quantities during that same time period in all other years during the period of record. These cumulative precipitation values over time in each grid cell are fit to a gamma distribution to account for non-normal precipitation distributions, and shape and skewness parameters are calculated. These parameters are then used to transform the gamma distribution to a Gaussian normal distribution. The SPI is defined as the number of standard deviations that the observed normalized cumulative rainfall deviates from the long-term mean, using a minimum 30-year record (Edwards and McKee 1997; Hayes 2000). Thus, the long-term mean is represented by an SPI of zero; and anomalously wet and dry periods are represented by positive and negative numbers, respectively. The durations over which the SPI can be calculated are flexible. The three-month SPI could compare the total precipitation in the three months up to and including August, for example, to the same three-month window for each year in the period of record. Figure 3 shows the running one-, three-, twelve-, and thirty-six-month SPIs at one example grid cell over the Horn of Africa.

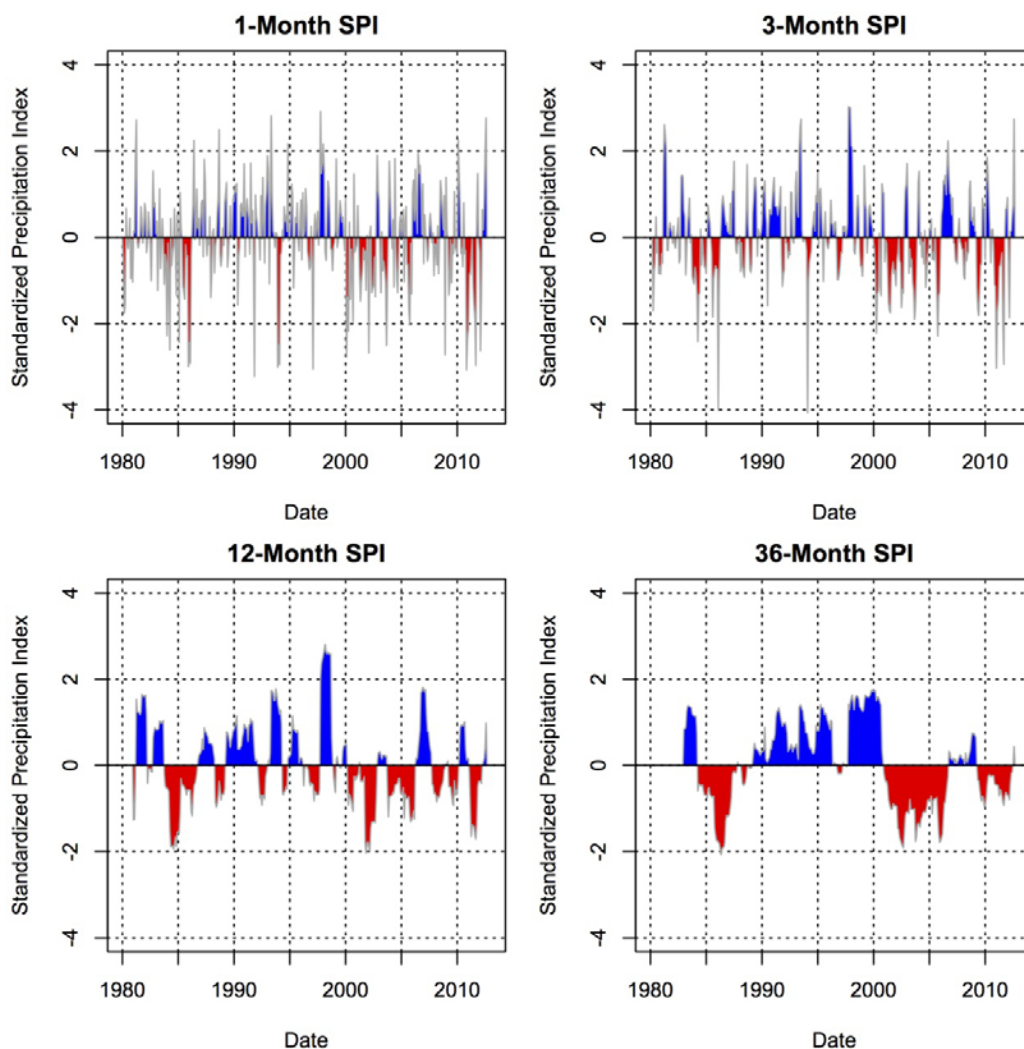


Figure 3. Four SPI time series from one grid cell location over the Horn of Africa. Red indicates below-average precipitation at that location over the period of record.

In this study, we calculated SPI indices over Somalia and Kenya from 1980 to 2012 using the aggregate precipitation from time periods of 1, 2, 3, 4, 6, 8, 12, 18, 36, and 48 months, yielding eleven separate indices, each with one value per month per grid cell in the time series. Functions from the *R* library *ncdf* were used to transfer MERRA-Land data into *R*-compatible data frames, and the *R* library *SPEI* was used to process the SPI time series. *R* libraries *sp* and *rgdal* were used to plot time-series spatial grids for the moisture index data frames. As an example, Figure 4 shows the gridded 3-month SPI series over Somalia from December 2010 through November 2011.

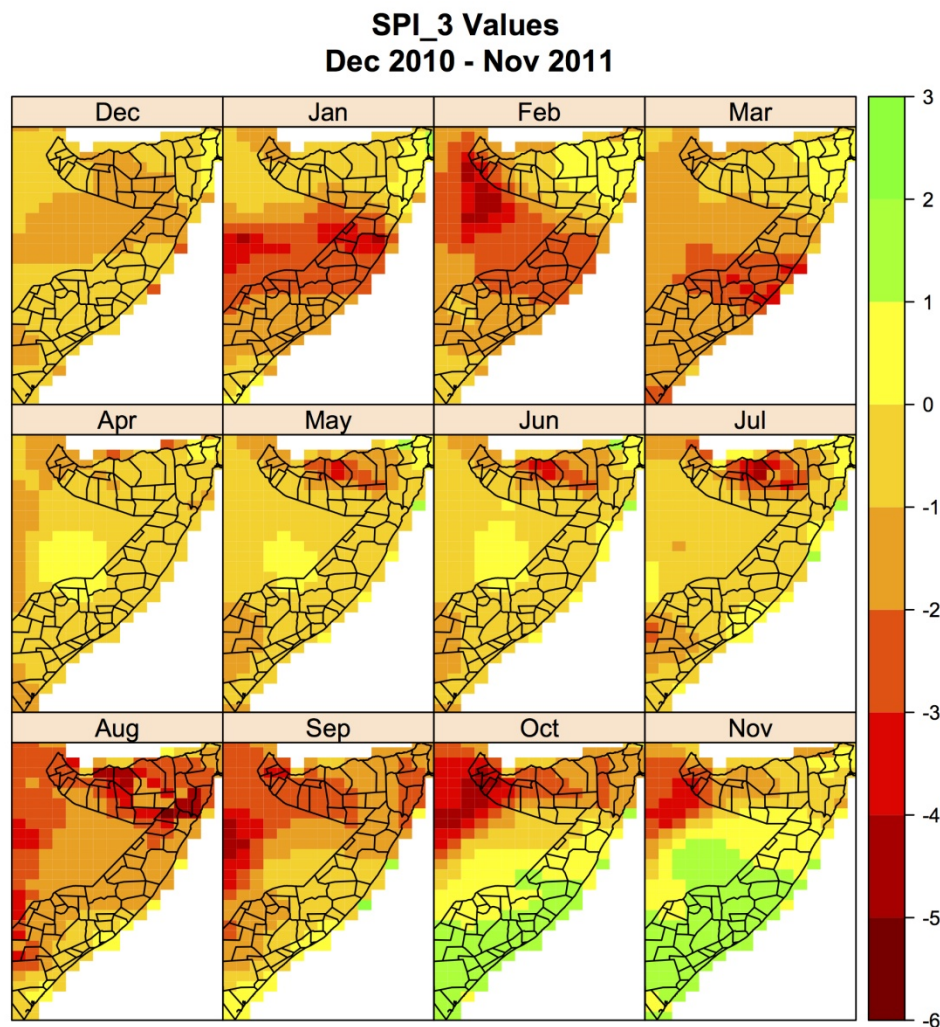


Figure 4. Three-month SPI values from December 2010 to November 2011, based on MERRA-Land precipitation data over Somalia from 1980 to 2012, with administrative districts outlined in black. Note that according to the MERRA-Land dataset, severe drought conditions were not limited to the southern part of the country and that the 2011 late-season minor rains (*Deyr*) were above normal in the south, contributing to the end of the drought.

2.2.2 Standardized soil water index

In principle, any scalar gridded model output or interpolated observational dataset with a sufficiently long historical record could be used as an input to a standardized index based on the SPI. In addition to the precipitation-based SPI, this study also used the MERRA-Land root zone soil wetness product to derive a Standardized Soil Wetness Index (SSWI). While an SPI may be a preferred indicator for periods of hydrologic drought, SSWI may be a better indicator of periods of agricultural or pastoral drought because the root zone soil moisture may be more closely tied to vegetation response to climate conditions than to rainfall alone. A simi-

lar product could be calculated for runoff or for any other modeled output with the code developed here.

2.2.3 Benefits of standardized indices

There are several potentially important benefits that standardized indices offer. First, they take into account seasonal differences in moisture distribution so that a typical dry season can be distinguished from an anomalous rainy season failure. Second, they allow for calculation of indices at a variety of temporal scales; this study calculated all moisture indices at 1, 2, 3, 4, 6, 8, 12, 18, 36, and 48 months. Social and economic responses to drought may, in a given location, depend on the temporal scale at which drought is defined; and standardized indices provide a way to incorporate this temporal scale into the analysis. For example, a family tending a herd may be minimally impacted by a short drought as they have other assets to sell or suffer only a partial loss of their herd; but as the drought continues and a second rainy season fails, their response changes to migration. We aim to use these indices at multiple temporal scales to ascertain, across the larger population, what the temporal thresholds are for such responses to drought. Third, standardized indices are not theoretically bounded, allowing for the inclusion of unprecedented values should new moisture extremes occur in the future. They provide a flexible framework for analyses and can be modified to create related indices using other model output variables or combinations thereof from any gridded or atmospheric land surface model or interpolated dataset. For example, the same analysis could also be used to assess flooding risk if data were input at the shorter time intervals typically associated with flood events (Seiler et al. 2002). Although not explicitly calculated in this study, standardized indices can also help define discrete droughts and quantify their onset, duration, and severity. These calculations can be based on, respectively, a threshold value of the index score, time to return to the long-term mean, and the total area between the index curve and zero during the drought (McKee et al. 1993). Therefore, development of a code base using these two drought indices (SPI, SSWI) is not solely an end in itself but could also serve as a platform for further research.

2.2.4 Additional considerations when using standardized indices

SPI-like indices are based on deviations from some historical norm, and incorporation of new data requires appropriate treatment. If climate datasets are subject to ongoing additions of new data or new data assimila-

tion techniques not available for the entire time period, a period of reference would have to be defined to maintain consistency in index values. In the case of spatial changes in resolution, the calculated historical norms could, therefore, change in any given location and would have to be recalculated. The current study uses the period from January 1980 to August 2012 as the period of reference, using monthly outputs from MERRA-Land as inputs to SPI-type calculations.

One important consideration is that, as used in this study, SPI and SSWI are technically moisture indices rather than drought indices alone. Negative values are indicators of drought-like conditions, and positive values are indicators of anomalously wet periods. In certain circumstances, the occurrence of floods and consequent socio-economic effects may interfere with the use of the SPI/SSWI as drought indicators; and it may be preferable to use only the negative values of the index that represent drought. In the study location and time period, we presumed any flooding to be localized in both time and space with respect to the widespread nature of the drought; and therefore, we did not apply in this analysis a reformulation of the more generic moisture index.

2.3 Drought and population movement

2.3.1 Population movement tracking data

In 2006, the United Nations High Commissioner for Refugees (UNHCR) began funding an effort to collect, verify, and share information on population movements affecting civilians inside of Somalia, even in areas where conditions of conflict and insecurity restrict access for staff of non-governmental organizations (NGOs). Prior to this time, little to no data were collected on IDPs within Somalia. Members of a network of 67 local non-governmental organizations (LNGOs) each submit a monthly standardized form to report recent counts of displaced people along with their locations, origin, and reason for movement. These data are collected via routine visits to established IDP settlements, bus stations, checkpoints, and ports and by speaking with village elders. The data are compiled, analyzed for duplicate information, verified by third parties as necessary, entered into a database, and shared with the international community via the Population Movement Trends (PMT) portal (UNHCR 2013). These records classify reasons for displacement into one of 11 categories: insecurity, drought, lack of livelihood, flood, clan conflict, IDP return, fire, forced return, eviction, relocation, or cross-border movement. Our study used both

the drought- and insecurity-induced population movement datasets, reported by the source district of the displacement, at monthly intervals from January 2008 to September 2012.

Raw data from the PMT dataset are given in numbers of people per district; however, we performed statistical analysis on this data as a proportion of the total district population. An area suffering from drought that saw 5000 people migrate during a given January would have different implications if that number was a very small fraction of the total district population versus if it constituted a large proportion of the population. Therefore, we recalculated PMT data as a proportion of the district's baseline population. Using population distribution models appropriate to country-specific data availability and geographic characteristics, Oak Ridge National Laboratory (ORNL) has produced the LandScan global population datasets (UT-Battelle LLC 2010) at a 30×30 seconds (approximately 1×1 km) resolution. The dataset is freely available to US government federal agencies through ORNL. We estimated the district population by using zonal statistics in ArcGIS on Somali district polygons overlying the LandScan 2010 dataset. Our study used the 2010 version of the dataset to estimate drought-induced internal displacements as a percentage of the ORNL-based district-level population count.

LandScan does not provide explicit error estimates for its products or specify the extent to which manual verification and modification have been applied in a given region, but it does apply corrections to its models using high resolution imagery "as time and budget constraints allow" (ORNL 2013). In the case of Somalia, where independently documented refugee flows crossing international borders have been significant, it is reasonable to allow that the LandScan dataset could be an important, though by no means the only, source of error in this analysis. However, in the absence of other more reliable census data for Somalia, it is also reasonable and desirable to achieve a first-order estimate of population movements as a fraction of some original, if estimated, district population.

2.3.2 Spatial resolution and alignment of gridded and polygon datasets

Combining two datasets for analysis that each have different spatial resolutions and characteristics requires application of a suitable and repeatable method. Below, we describe two methods used in different aspects of this study.

2.3.2.1 Majority area

The resolution and location of gridded MERRA-Land products required a method of assigning grid cells to the geographic units of interest in a given statistical analysis. In the case of the PMT datasets, these units of interest are Somali administrative units at the district level (Fig. 5). Because the spatial resolutions of these grid cells are generally similar to those of Somali administrative districts, we assigned grid cells to a specific district when at least half of the area of the grid cell fell within that district's boundary. Subsequently, we manually assigned one or more grid cells that best represented the spatial extent of the district to the seven districts that were not large enough to intersect more than half of any one grid cell. The result of this method of assignment, implemented using the *over* function in *R*'s *sp* package, was that each district was represented by between one and seven grid cells and that twelve grid cells represented more than one district.

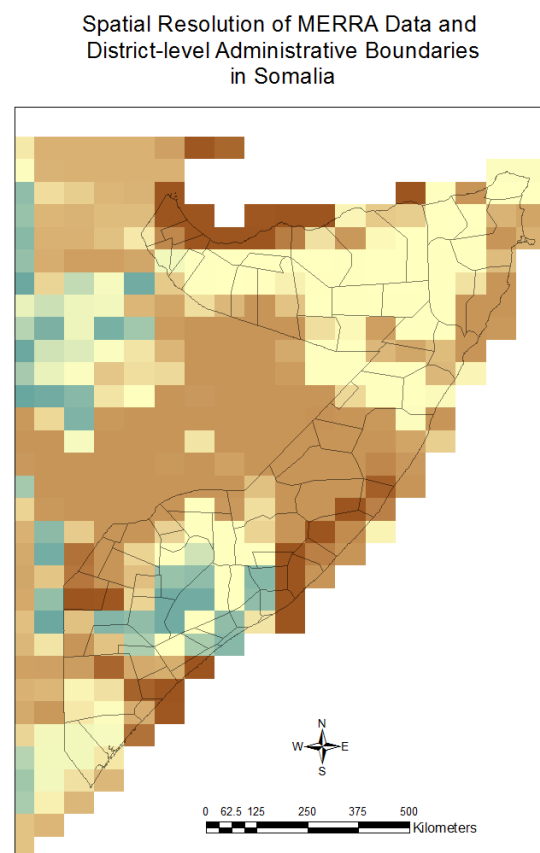


Figure 5. MERRA-Land grid cells are displayed together with Somali administrative districts (black lines). Using the majority area method, each Somali administrative district was associated with between one and seven MERRA grid cells, and some grid cells were associated with multiple districts.

2.3.2.2 Dense hexagonal point sampling

In other datasets, such as the livelihood zone datasets, polygons were often small and irregularly shaped with respect to the grid cells, particularly in livelihood zones associated with rivers. In these cases, we used a dense hexagonally-distributed point sampling method where we sampled points over the irregular polygon layer. We used the coordinates of the sampled points to select drought indices from the MERRA-Land grid cells in which those coordinates fell. We then calculated statistics, such as mean and standard deviation of the drought indices, within a polygon. At fine enough sampling density (e.g., 10,000 points throughout the country of Somalia), this method intends to approximate the effect of an area-weighted sample. However, it does not overcome the fundamental limits of the grid cell resolution with respect to accurately representing climate conditions in small and irregularly-shaped polygons.

2.3.3 Correlation methods

Both SPI/SSWI moisture indices and the PMT datasets provide time series data in the form of continuous variables. We used correlation coefficients to measure the strength of the associations between the various moisture indices at each grid cell and the PMT in the districts to which the grid cell had been assigned by the majority area method. We explored two methods of calculating correlation with a goal of identifying the moisture index (or indices) that might be the best predictor of population movements in situations similar to the 2011 Horn of Africa drought.

2.3.3.1 Pearson's r

Pearson's r measures the linear association between two continuous variables. Specifically, it normalizes the variation of each data point from the mean over the standard deviation and effectively averages the products of all paired normalized values over the dataset (x_i, y_i) :

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

where

n = number of sample pairs

s = standard deviation of the dataset.

Pearson's r presumes that the paired data follow a bivariate normal distribution and is strongly influenced by outliers. While moisture indices have by definition been transformed to adhere to a normal distribution as part of the moisture index derivation, drought-induced population movement data are typically skewed with frequent values at or close to zero and sporadic higher values as a population responds to more severe drought conditions. However, to the extent that in this case it is precisely the extremes in this population movement that are of primary interest, Pearson's r may be a useful statistic despite the fact that one of the datasets is generally not normally distributed. Typically, misapplication of Pearson's r results in reduced sensitivity to real effects, and it therefore can be expected to underestimate rather than exaggerate results (Helsel and Hirsch 2002; Bradley 1968). Pearson's r was therefore one method used to measure the association between moisture indices and PMT data.

2.3.3.2 Kendall's τ

Kendall's τ is a measure of the strength of the association between the ranks, rather than the values, of two paired datasets (x_i, y_i) . It therefore lessens the impact of outliers and variables that exhibit skewness. It is worth noting that values of τ , though they may also vary from -1 to 1 , are not equivalent to values of Pearson's r on the same scale; a τ of 0.7 is considered to indicate as strong a correlation as an r of 0.9 , for example (Helsel and Hirsch 2002).

Tau is computed by ordering the data pairs (x_i, y_i) by increasing x . Then, each y is compared to every other y associated with subsequent, higher-ranked x values. A count is made of "concordant" pairs, where the y associated with the higher ranked x value is higher than that associated with the lower ranked x value, and "discordant" pairs, where the reverse is true. The difference between concordant and discordant pairs is then divided by the total number of comparisons among the data pairs. Kendall's τ is then given by

$$\tau = \frac{S}{n(n-1)/2}$$

where

S = the difference between the number of concordant and discordant pairs

n = the number of data pairs.

The PMT datasets have large numbers of tied values at zero representing no population movement, and these and other tied values were counted neither as concordant nor discordant.

2.3.3.3 Significance levels

After a test statistic r or τ is calculated, the significance of that statistic must also be calculated to determine whether the statistic is significantly different from what would be expected were a null hypothesis true (“no correlation between the datasets”). A statistical metric for significance is the p -value, which indicates the probability that the test statistic could be obtained as the result of chance alone and which depends in part on the number of data points used to calculate the test statistic. This study used two-sided significance tests with a significance level of 95%. Although one may expect that IDP movement would increase following periods of negative moisture indices and, thus, it may be justifiable to set up a one-sided hypothesis test to look specifically for negative correlation coefficients, it would also be true that if a region exhibited a so-called drought-induced migration response during a period of abundant moisture, that result would also be of interest, either because it could indicate a problem with the drought index in that region; a problem with the PMT dataset; or possibly a response to some external action that may be associated with drought, such as the provision of aid. Therefore, we calculated a two-sided p -value for each correlation coefficient and deemed p -values less than 0.05 to be sufficiently significant as to reject the null hypothesis that there is no correlation between the PMT and moisture index in that location. We then mapped both correlation constants and p -values for subsequent qualitative spatial analyses.

Where ties are present in the data as occurs when multiple months have zero population movement, calculating the significance of Kendall’s τ necessitated a special treatment of the total number of rankable comparisons among the data pairs. A correction is applied to the significance calculations to account for the number of ties and the number of values involved in each tie in the dataset. This correction applies to a large-sample approximation only; and therefore Kendall’s τ , when ties are present, should only be calculated for datasets where $n > 10$ (Helsel and Hirsch 2002). To meet this condition in the context of this study, correlation tests using Kendall’s τ were calculated only for time periods of one year or longer. Corrections

for ties were calculated with the *R* package *Kendall* (McLeod 2011). Kendall's τ , with significance levels corrected for ties, offers the most robust correlation analysis because of the frequency of ties of zero in the PMT dataset. However, we note that the data pairs of most interest occur not among these tied values but where population movements are large; these are the values emphasized by Pearson correlation coefficients. Therefore, we suggest that both Pearson and Kendall correlation methods are reasonable for use in this study; and we present the results of both methods.

2.3.4 Livelihood zones

Based on the ways in which people earn their livelihoods, one could hypothesize that certain populations would be more likely to relocate during a drought than other populations. It is therefore of interest to assess whether the correlations between drought indices and drought-induced population movements might differ across subpopulations. As an activity of USAID, FEWSnet is responsible for developing a framework to understand and predict emerging food security issues in parts of the developing world. It produces livelihood zone maps that summarize information on household income sources relevant to food security issues and are designed to help predict the potential impacts of various types of shocks. In our study, correlations between district-level drought-induced population movements and moisture indices were aggregated via area-weighted sampling by the 34 FEWSnet livelihood zones identified for Somalia (Fig. 6). This aggregation of results by livelihood zones allowed us to assess whether certain livelihood zones or livelihood zone types had been empirically more susceptible to internal displacement in response to drought between 2008 and 2012. The analysis was run on all results (both Pearson and Kendall correlation methods, for SPI and SSWI, and at all eleven time scales). We used livelihood shapefiles for Somalia, updated in 2008, in our study (FEWSnet 2011). Because the scale and shape of identified livelihood zones are substantially more varied than that of districts or grid cells, we used the dense hexagonal point sampling technique to approximate area-weighted sampling.

Livelihood Zones in Somalia

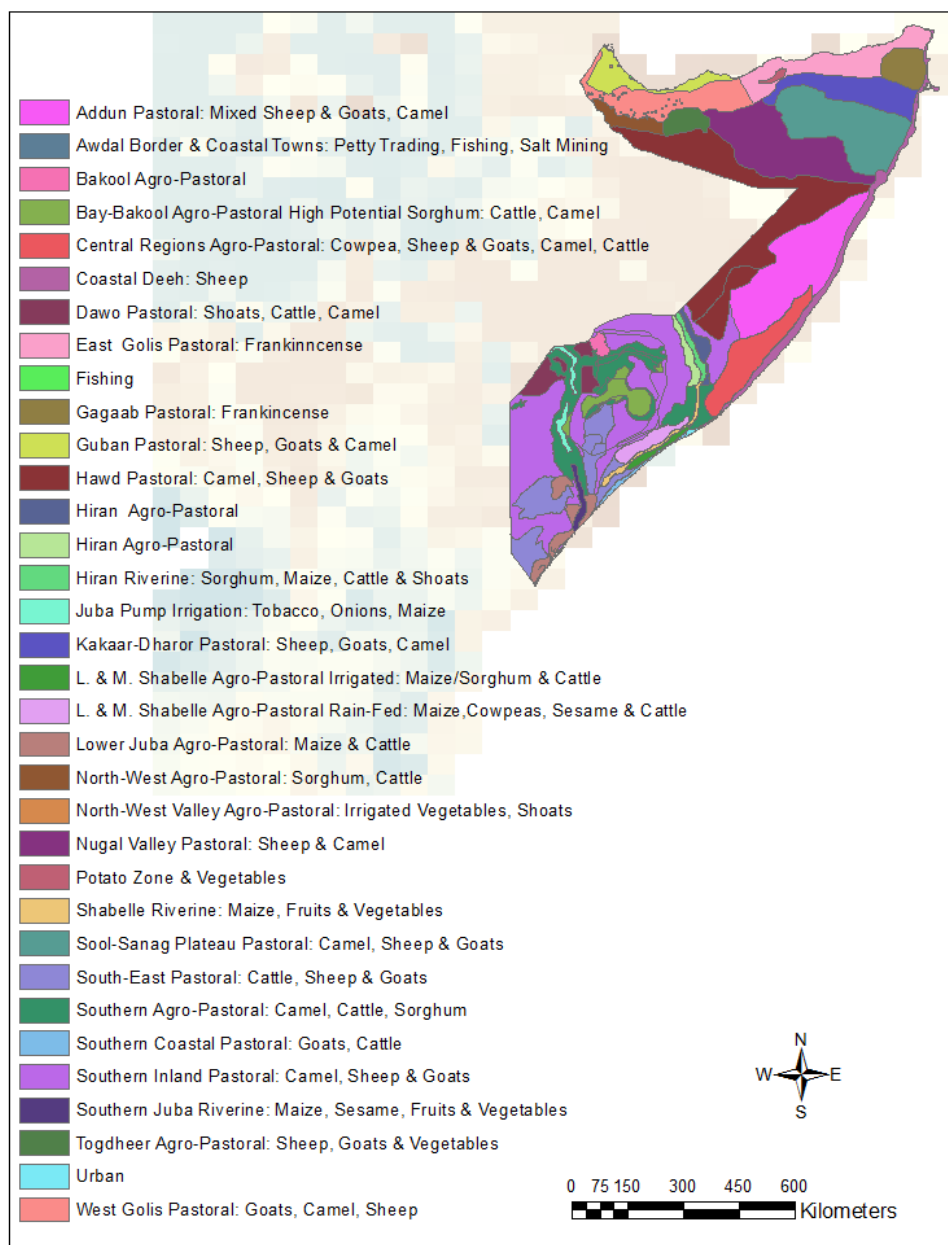


Figure 6. Livelihood zones in Somalia with MERRA-Land grid cells in the background.

2.3.5 Linear models

Could there be a relatively simple way to use livelihood zone information in conjunction with moisture indices to understand and predict population movements that occur due to drought? In the analysis up to now, assessing correlations between moisture indices and movements over time has given as much weight to anomalously dry months during dry seasons as to anomalously dry months during rainy seasons. In reality, however, a

“rainy season drought” would be expected to have a much stronger socio-economic effect than a similar “dry season drought.” Therefore, finding a method to weight drought indices based on the time of the year in which precipitation is most important to support livelihoods in a given area might provide support to a predictive model. In addition to livelihood zones, FEWSnet also produces country-level seasonal calendars, indicating the times of year of primary importance to various aspects of the agricultural and economic cycle (i.e., principal growing seasons, harvests, and livestock kidding and lambing).

We used the Somali seasonal calendar (Fig. 7) (FEWSnet 2013) to weight the importance of different months of the year within different livelihood zones. Based on the seasonal calendar, important months were assigned to each livelihood zone as follows. Agro-pastoral zones were presumed to depend on all months during the two typical rainy seasons: April–June and September–November. Pastoral livelihood zones were presumed to be dependent primarily on the kidding and lambing months of June and October, presuming a certain degree of population mobility during other seasons that, unlike agricultural-based livelihoods, would allow for movement of assets to more suitable climatic areas. Both riverine and irrigated agricultural areas were, for the present purposes, considered to rely equally on all months of the year. Coastal and urban livelihood zones were also presumed to rely on all months equally as the principal livelihoods in these locations are less directly dependent on agricultural or pastoral cycles.

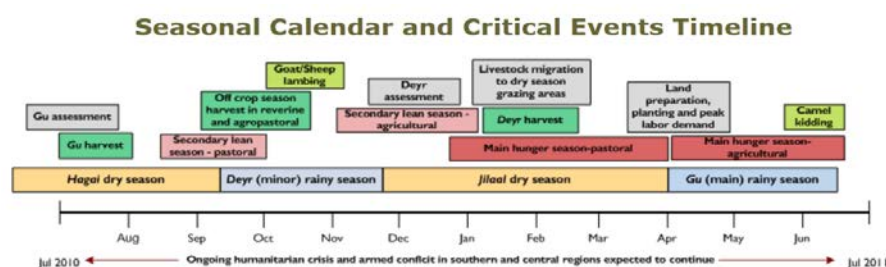


Figure 7. FEWSnet seasonal calendar for Somalia.

Each district was then assigned a set of weights for each month of the year based on the areal percent of each livelihood zone type inside that district. We multiplied the moisture indices throughout the year of interest by the weight of the respective month in which they occurred; and the sum of these results over the year constituted a new, weighted moisture index with which to test the hypothesis that drought-induced migration would be a linear function of this new livelihood-weighted moisture index. In

particular, a negative slope of the linear function would be indicative of increased population movement with more strongly negative values of this weighted moisture index.

We then compared this linear model to that of a simpler, unweighted drought index over the same time period to ascertain whether incorporating livelihood zones and the seasonal calendar could help distinguish population response. We also divided the data into geographical subsets to assess the effects of certain governance factors in the linear models, and we mapped model residuals for further analysis. To assess which districts might be considered outliers in the model, we used graphs of residuals versus fitted data, normal quantile-quantile plots, and residuals versus leverage. Subsequently, we ran both linear models in “predictive” mode to provide estimates of population movements based on the linear model along with the confidence limits of the models.

2.4 Drought and market prices

2.4.1 Market price data

The Food and Agriculture Organization (FAO) of the UN’s FSNAU partnered with FEWSnet to monitor on a monthly basis market prices of 33 different market items in Somalia, producing a dataset beginning in 1995 that includes food, fuel, daily labor rates, exchange rates, and construction and household goods (Table 1). The dataset covers up to 44 market locations throughout Somalia though coverage of all items is not complete at all market locations. Data used in our study was obtained from FSNAU’s Integrated Database System (FSNAU 2013).

Table 1. Market products monitored by FSNAU (2013).

Market Products
CAMEL, LOCAL QUALITY
CATTLE, EXPORT QUALITY
CATTLE, LOCAL QUALITY
CEMENT, 50KG
CHARCOAL, 50KG
COOKING POT, ALUMINIUM 7L
COWPEAS
DAILY LABOR RATE
DIESEL, 1 LITRE
GALVANISED IRON SHEET, GA 26
GOAT, EXPORT QUALITY
HOLLOW CONCRETE BLOCK, 10CM X 20CM X 40CM
IMPORTED RED RICE, 1KG
LOCAL SESAME OIL, 1 LITRE
NON COLLAPSABLE JERRYCAN, 10L
PETROL, 1 LITRE
PLASTIC TARPULIN, 14 X 5 METRES
RED SORGHUM, 50KG
ROOFING NAILS, 15KG
SHEEP, EXPORT QUALITY
SOMALI SHILLING TO KSH
SOMALI SHILLING TO USD
SOMALILAND SHILLING TO DJIBOUTI FRANC
SOMALILAND SHILLING TO ETHIOPIAN BIR
SOMALILAND SHILLING TO USD
TIMBER, 2IN X 4IN X 20FT
WATER DRUM
WHEAT GRAIN, 1KG
WHITE MAIZE, 50KG
WHITE SORGHUM, 50KG
WOVEN DRY RAISED BLANKET, 150CM X 200CM
YELLOW MAIZE, 1KG

KHS = Kenyan Shilling

USD = US Dollar

2.4.2 Market price ratio to moisture index correlations

In the same way that we calculated and mapped correlation constants to permit visualization of the relationship between population movements to MERRA-Land-derived moisture indices, we also mapped correlations between market prices and moisture indices. We used the geographic coordinates of a single point for each market, evaluated all correlations between the market data and the moisture index data at only a single grid cell at each market location, and mapped the results at the point location of the market. Rather than using raw market prices, we used a ratio of market prices to the local day labor rate to represent local purchasing power for market goods. To make this process easily extensible to other locations in the future, we developed code that was able to automatically extract the approximate coordinates of 42 of the 44 markets based only on the name, province, and district of each market, using a series of calls to the online GeoNames database (GeoNames, n.d.).

2.4.3 Cross-correlation functions

How long does it take for the effects of a drought to be reflected in market behavior? Coincident to or following a drought, one expects that as certain agricultural yields diminish, prices for some products in local markets will rise. At the same time, drought can force pastoralists to cull their herds, leading to a glut in local markets and decreased prices. It is also expected that as agricultural work opportunities decline with drought, the day labor rate, which represents the ability of an individual to bring in income in the absence of other assets, will decrease in agricultural areas. The time lag of these price-to-day-labor-rate ratios following drought may vary from location to location and with local meteorological and economic conditions. Using this market price data to extract this information may provide a measure of which subpopulations in the area may have been affected by drought and may help establish typical time scales for social stress following this type of climate event.

An autocorrelation function can be used to measure the degree to which a time series function is correlated with itself at different time lags. The autocorrelation for a time series function X of length n and lag h is defined as

$$\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}$$

where

$$\hat{\gamma}_X(h) = \frac{1}{n} \sum_{i=1}^{n-h} (X_i - \bar{X})(X_{i+h} - \bar{X}).$$

By definition, a function is perfectly correlated with itself ($\hat{\rho} = 1$) at a lag of zero.

If, rather than analyzing the correlation of one function with itself, one desires the correlation of that function with some other function at different lag times, then the process is termed cross-correlation. In our analysis, the ratio of a product's market price to the day labor rate is used as one time series of interest. Then, these price ratio time series were cross-correlated with moisture index time series to determine if there were any particular combinations of products and moisture indices for which certain lag times are predominant. We then mapped results by market location to determine if there were any spatial patterns of interest in these lag times. Cross-correlation functions (CCACFs) were calculated using the *stats* package in *R* (R Core Team 2012a) with a maximum lag value based on the number of elements, n , in the time series and the number of different time series, m , used in the analysis, such that

$$\text{maximum lag} = 10 \log \left(\frac{n}{m} \right).$$

Initially, we calculated CCACFs for all time series, regardless of whether or not they contained periods of missing market price data. For products in which these periods were significant, it was necessary to choose subsets of the record for which values were most complete. In the current analysis, this was accomplished by limiting the datasets to a shorter 12-month time series from December 2010 to November 2011 and excluding analysis of products at market locations that lacked complete data coverage during this time period. We calculated confidence intervals for each lag time of 95% by using $\pm 1.975\sqrt{n}$ (R Core Team 2012a) and plotted on a map values with this confidence level or higher. In addition, if there was more than one lag time for which autocorrelation was significant, only the most significant lag time was displayed on the map.

2.4.4 Network analysis

Whereas the two techniques above calculate certain market-climate characteristics (correlation and lag times) for each location, network analytic techniques can aid in attempting to understand other spatial relationships in the market data. To the extent that time series similarity indicates some measure of market connectivity, do patterns of market connections change during drought years? Are there groups of prices (either by location, by product, or by any other classification) that behave similarly? The analyses that follow examine monthly market price data by using several techniques from network analysis to understand market response previous to and during the December 2010–November 2011 period when drought-caused population movements peaked.

2.4.4.1 *Inter-market price correlations*

We calculated Pearson correlation coefficients for pairs of market price time series for each possible pairing of the 44 markets in the dataset and for each of the 33 FSNAU market products. The correlations were calculated for the drought year (December 2010–November 2011) and for the two preceding, non-drought years. In a price correlation matrix, we retained correlation constants that were significantly different from 0 at the 95% confidence level. Further, these values were filtered with a hard threshold whereby only values greater than 0.5 were retained in an adjacency matrix. Each matrix was then transformed to a network graph. The nodes of the graph represented individual markets, and the edges of the graph represented the degree to which prices for a particular product were correlated between the two markets linked by the edge. For the correlation test to correctly calculate p -values, rare negative correlation values were removed from the dataset.

2.4.4.2 *Degree centrality*

Degree centrality is the number of edges connected to each node in a graph and is one measure of how well connected a particular network node is to other nodes in the network. Our study used degree centrality as a measure of how well correlated the price of a market product was to other prices in the country, with the intention to use this data to identify markets or products that may be either particularly economically connected to or isolated from other markets. We used the same correlation and significance methods as for the inter-market price correlations above and ob-

tained results both for individual products and for the summed degree centrality at a given market across all products. We calculated these values for the December 2010–November 2011 drought year and for the two preceding years to assess how the drought may have impacted connectivity of different markets.

2.5 Census data and vulnerability to drought

The University of Minnesota’s Minnesota Population Center has developed the Integrated Public Use Microdata Series (IPUMS) International, which provides open access to select international census data for use in social and economic research. From publically available census data for Kenya from 1989 and 1999, we extracted district-level data related to household water infrastructure (Minnesota Population Center 2011). The census data constituted a 5% sample of the total number of households and totaled approximately 2.5 million census records. A more recent census was not yet publically available at the time of this study.

Out of the 179 variables available in the IPUMS datasets for Kenya, we chose a small selection of household-level variables to characterize household access to water infrastructure (Table 2). In northern Kenya in particular, where nomadism and transhumance are common, this data could differentiate levels of sedentarization and potentially, therefore, be an indicator of short-term population stability during drought. Variables characterizing ownership status, which are not directly related to water infrastructure but may also be related to population stability, were also selected. Roofing material was included in this analysis for its potential use as an observable variable that may be assessed remotely and provide links to the other variables.

Table 2. Select Kenyan household census variables.

Census Variable {Variable Name}	Values	{Value Name}
Urban/rural categorization {URBAN}	Rural	{rural}
	Urban	{urban}
Census district {DISTKE}	87 census districts were mapped to 48 administrative districts. IPUMS does not currently provide shapefiles of the respective census districts.	{<DistrictName>}
Ownership status of the housing unit by some member of the household {OWNRSHP}	Owned	{own}
	Not owned	{rent}
Detailed ownership status {OWNRSHPD}	Owned, already paid (including acquisition via mortgage)	{owned.pd}
	Owned, constructed	{owned.const}
	Owned, inherited	{owned.inh}
	Renting from government	{rent.gov}
	Renting from local authority	{rent.la}
	Renting from parastatal authority	{rent.ps}
	Renting from private company	{rent.pc}
	Renting from individual	{rent.ind}
Physical means of household water supply {WATSUP}	Piped water	{piped}
	No piped water	{no.piped}
Sewage disposal {SEWAGE}	Sewage system (public sewage disposal)	{sew.pub}
	Septic tank (private sewage disposal)	{sew.septic}
	Not connected to sewage disposal system	{no.sew}
Access to toilet {TOILET}	No toilet	{no.toilet}
	Flush toilet	{flush.toilet}
	Non-flush, latrine	{latrine}
	Other toilet	{toilet.other}
Predominant roofing material {ROOF}	Concrete or cement	{roof.conc}
	Tile	{roof.tile}
	Sheet metal	{roof.shtmtl}
	Zinc or tin	{roof.zc.sn}
	Grass	{roof.grass}
	Palm or makuti	{roof.palm}
	Asbestos	{roof.asb}
	Other	{roof.other}

2.5.1 Mapping of census data

As a first step in data exploration, we mapped the variables in the 1989 dataset. At the time of the study, neither IPUMS nor the Kenyan National Bureau of Statistics made available shapefiles outlining the geospatial census units in which the data were collected. Inspection of names of census units and maps of Kenya indicated that many units appeared to correspond to administrative districts. However, some census units were not equivalent to official administrative districts, with census data provided at finer scales than that of the districts. In other cases, sub-district samples appeared to have been aggregated to the district level for public dissemination. Several census units appeared to overlap or cross administrative district boundaries or were otherwise unclear, and these districts were not included in the mapping. Census data was mapped in *R* using code from the library *UScensus2000* (Almquist 2010).

2.5.2 Association rules analysis

Association rule analysis (Agrawal et al. 1993), also known as “market basket analysis,” is a technique we used to link observable census variables to other unobservable variables. This allowed us to better characterize a community and to create more detailed analyses. For example, if we know that a rural community in a particular drought-affected district has sheet metal roofs, we could create a vulnerability analysis by querying the percentage of that type of household that has piped water or sewage systems, which could in turn be an indicator of how vulnerable those households may be to drought conditions. Because the level of spatial disaggregation of the Kenyan census data available to international researchers is limited, association rules analysis can help assess the mutual frequency of variables of interest, thus providing an avenue for disaggregation of census results in specific locations where some other external information is available.

2.5.2.1 Transactions and rules

In association rule analysis, *I* is a set of items or attributes and is called a *transaction* after the way the technique is applied in business analytics. For example, the set *I* may correspond to specific items (i.e., cereal, milk, bananas) purchased in a particular transaction. When using census data, the transaction represents a set of household-level attributes for one household derived from census questions (i.e., rural, does not have piped

water, has a sheet metal roof). A census database for a given year consists of the set of all sampled transactions, or households, together with the attributes of each household. The dataset is then mined for rules that associate certain “antecedent” items, X , with other “consequent” items, Y , such that both of the following are true:

$$X \rightarrow Y$$

$$X \cap Y = \emptyset$$

2.5.2.2 Measures of significance

Of the set of all possible rules that apply to a dataset, several measures of significance can assist in ranking the importance of the rules. The *support* of a rule indicates the proportion of the transactions in the database that contain the antecedent variables in X . For example, for the rule “if cereal and milk, then bananas,” the support would indicate the fraction of the total transactions that contained both cereal and milk. Support for X is denoted $supp(X)$. The *confidence* is the ratio of the support for both the antecedent and consequent sides of the rule together to the support for the antecedent side alone (Hipp et al. 2000).

$$conf(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)}$$

The confidence estimates the probability that a household with the set of attributes X also has the attribute Y : What fraction of those who buy cereal and milk also buy bananas? The *lift* is a third measure of interest that describes the confidence of a rule with respect to the support of its consequent side (Brin et al. 1997). The higher the lift, the fewer association rules have Y as a consequent item without also having X as an antecedent set of items. Lift, therefore, is an indicator of the strength of the particular association rule as compared to other possible rules for the same consequent item in Y : How likely is it that a transaction that includes bananas means that cereal and milk have also been purchased?

$$lift(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X) \cdot supp(Y)}$$

2.5.2.3 Data preparation

In this study, the raw census data required some restructuring for association rules analysis. We pared down the data to reflect the most recent census year (1999) and one household per transaction, eliminating repeat transactions for individual household members. Invariant, unique, and redundant columns serving as census metadata were removed as were broader categories (e.g., ownership status and province) for which more detail was available in other variables (e.g., detailed ownership status and district). We then recast the categorical census data as a binary incidence matrix with one column per census attribute and one row per household, populating them with ones and zeros to indicate whether or not a given household was described by a given attribute.

We used the Apriori algorithm (Agrawal and Srikant 1994) to generate lists of association rules for the most recent census year. The Apriori algorithm is one of the faster algorithms for finding frequent itemsets (Borgelt 2003) and was implemented in the *R* package *arules* (Hahsler et al. 2012). We set minimum values for support and confidence and a maximum number of total items in a rule to generate each list. Resulting lists of rules using this dataset could reasonably consist of thousands to hundreds of thousands of entries. Each list could then be subset by known household characteristics (e.g., households in a given district with a certain roof type); and rules could be ordered by support, confidence, or lift for inspection. Both the rule generation and the rule inspection steps were iterated to explore specific examples of questions related to household water infrastructure in areas affected by the 2011 Horn of Africa drought.

3 Results and Discussion

3.1 Overview

Section 3 presents the results of the analyses for each dataset and method described in Section 2. It also discusses the potential utility of these results for understanding social and economic responses to drought and provides recommendations for further study. In many cases, the resulting datasets were too extensive to reproduce in their entirety in this report. In those cases, we present and discuss summaries, representative results, or results for select variables with the highest relevance to drought conditions. Section 3.2 presents representative results of the SPI and SSWI index calculations, and Section 3.3 provides the results of the methods used to analyze population movements in light of these indices. Section 3.4 presents results of the market price analyses, and Section 3.5 presents representative results of the techniques for census data analysis.

3.2 SPI and SSWI indices

We generated SPI and SSWI indices for the 1980–2012 study period over the entire areas of both Kenya and Somalia and for large parts of Ethiopia at time scales of 1, 2, 3, 4, 6, 8, 12, 18, 36, and 48 months. Figure 8 and Figure 9 present graphically two examples of these indices (1-month SPI and 8-month SSWI) over Somalia and eastern Ethiopia from January 2000 to August 2012. Other maps can be readily generated from the code, and subsequent statistical analyses used all resulting indices.

A comparison of these two indices illustrates the nature of the differences between these measures. Broadly, the 1-month SPI shows indices that change rapidly from month to month (Fig. 8) while the 8-month SSWI (Fig. 9) shows the slower progression of an index that incorporates data from the previous 8 months into each month's index. It is important to reiterate that the SPI is not equivalent to a map of monthly precipitation but that it rather compares monthly precipitation in a given set of months against all other years of record during that same period.

2000-2012: 1-month SPI

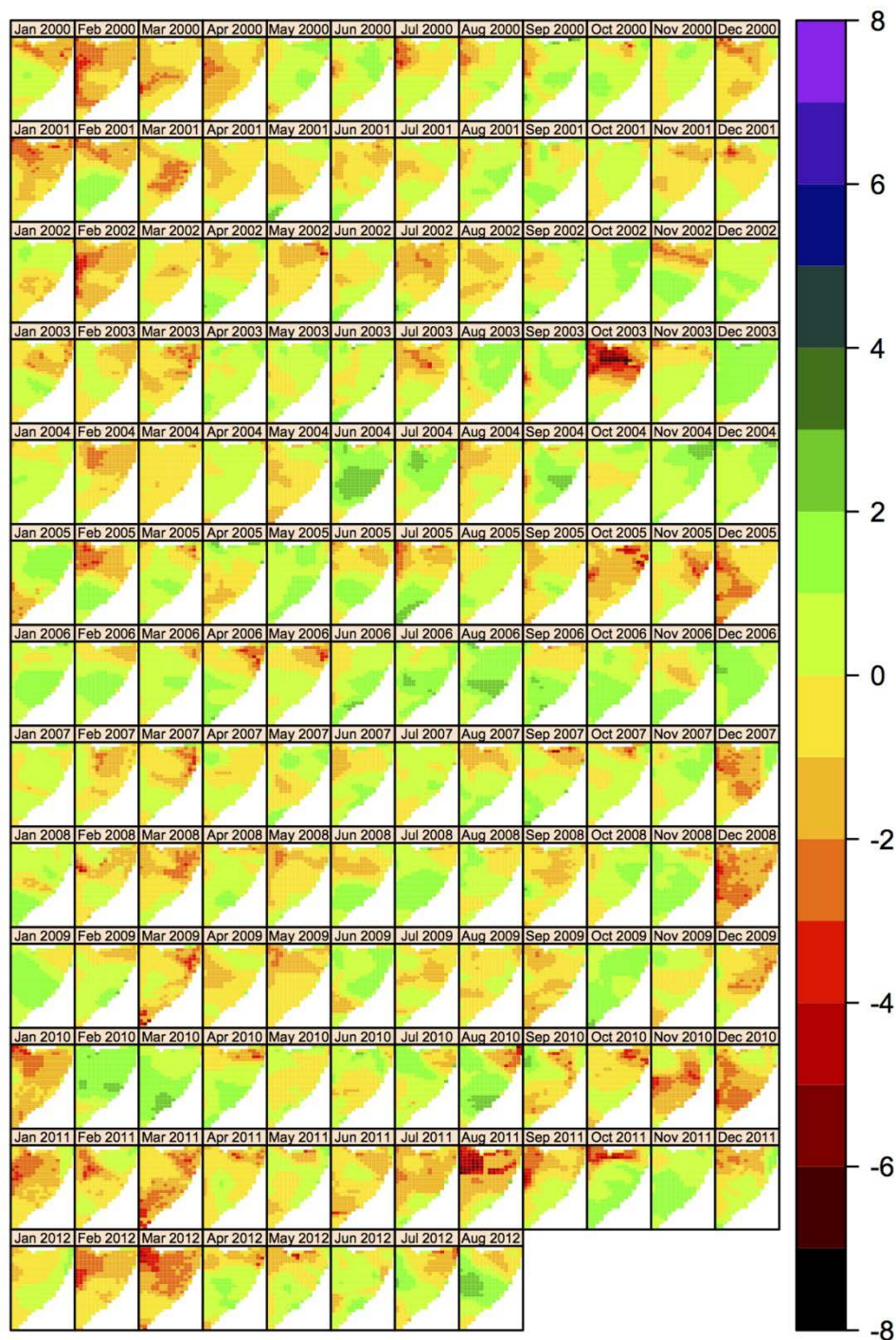


Figure 8. The 1-month Standardized Precipitation Index changes rapidly with changing precipitation values.

2000-2012: 8-month SSWI

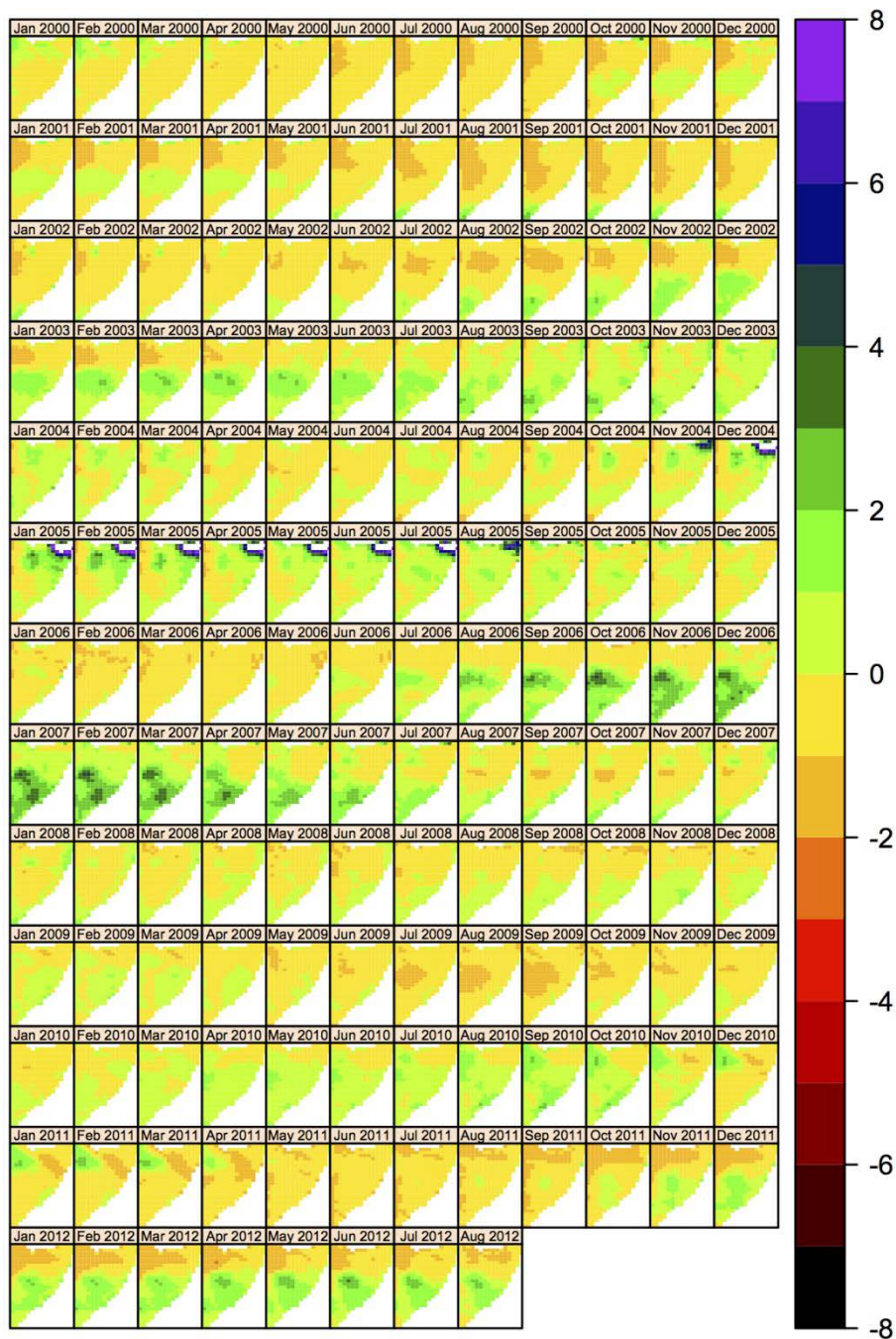


Figure 9. The 8-month Standardized Soil Water Index quantifies longer-term seasonal anomalies in modeled root zone soil moisture.

Precipitation and root zone soil values are related mathematically via the Catchment “Fortuna” v2.5 land surface model, but this relationship is not obvious between the two indices depicted here. Note, for example, how an unusually low SPI in northern Somalia during October 2003 has only a

mented effect on the 8-month SSWI for that month. This is because the input to SPI is routed as a raw precipitation value through the Catchment land surface model to the root zone and then blended with eight previous months' root-zone soil water data. Likewise, extreme values are also present in the 8-month SSWI values that are not matched by obvious features in the SPI dataset, for example, from November 2004 through August 2005 in the northeastern tip of Somalia. By using the MERRA-Land dataset to calculate moisture indices, any model parameters and weather conditions in the land surface model that created such wet root-zone conditions following a high but not extreme value of SPI in the area in November 2004 are implicitly accepted as a best-estimate form of "ground truth" and are not analyzed further in the context of this study. Qualitative comparison with other potential drought indicators would provide interesting comparisons for validation of these land surface models in future work. For example, normalized differential vegetation index (NDVI) anomalies from Moderate Resolution Imaging Spectroradiometer (MODIS)-derived imagery from a particular month or rainfall anomalies from the Tropical Rainfall Measuring Mission (TRMM)-derived datasets (Dinku et al. 2007) could potentially provide a way to evaluate the accuracy of the MERRA-Land precipitation data or derived moisture indices in the study region, despite the shorter period of record of these remotely-sensed products.

3.3 Correlations between drought-induced IDPs and moisture indices

The purpose of this phase of the study was to understand which drought indices and which geographic areas were most associated with drought-induced internal population movements over the duration of the PMT dataset (2008–2012). If significant relationships could be established based on these model-derived moisture indices, then it may be possible to anticipate the magnitude of future drought effects on similar populations. Results could also provide a baseline for future multi-country or cross-cultural population movement studies.

3.3.1 Overview and discussion of PMT data

The PMT dataset for drought-induced internal displacements from January 2008 through October 2012 is displayed in Figure 10, where the colored districts show raw out-migration numbers. The sampling techniques used to create this dataset did not track individuals over time and, there-

fore, may include multiple instances of out-migration of individuals or whole communities if they had returned in an interim period. Figure 10 displays the districts that people left and not the destination districts although that data is also available and could be incorporated in future analyses. The massive scale of population movements from December 2010 to November 2011 differs markedly from apparently more sporadic movements during other months in the dataset.

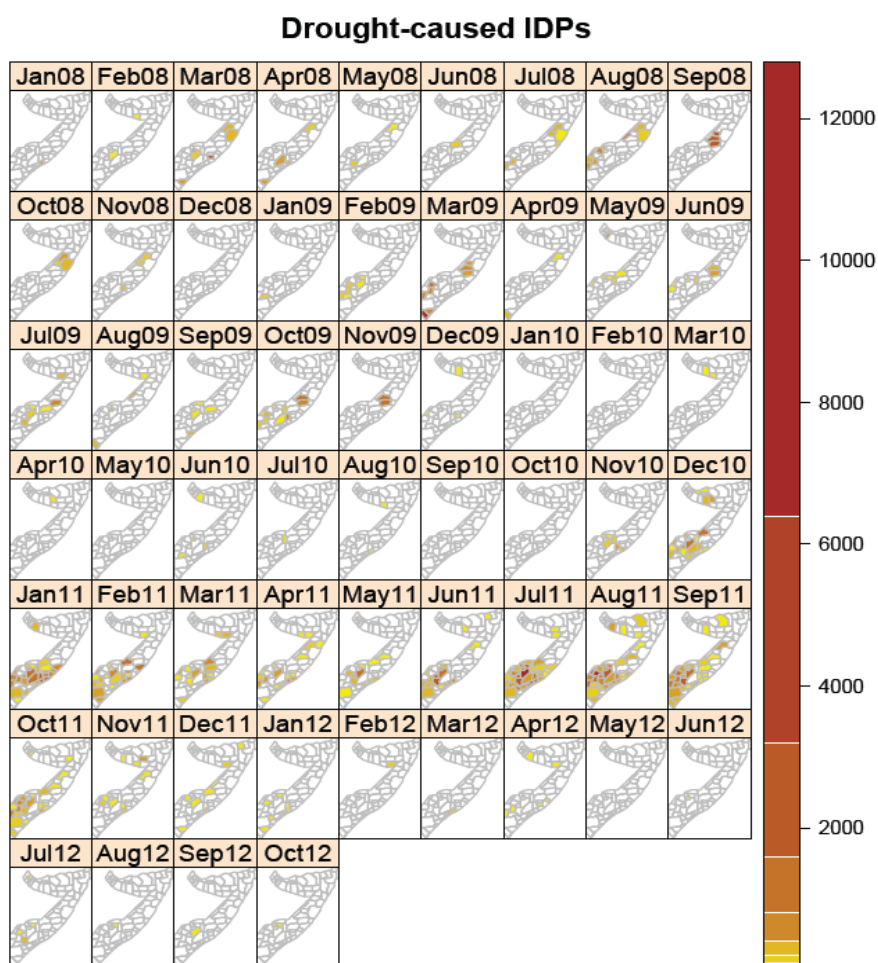


Figure 10. Numbers of drought-induced internally displaced people, by district of out-migration.

Analysis of insecurity-induced IDPs was not part of the analysis in this study but by way of comparison, the raw data are displayed in Figure 11. Note that the scales in Figures 10 and 11 have been stretched for optimal display and that the raw numbers for insecurity-cause IDPs are approximately an order of magnitude higher for the same color on the map as the drought-induced IDP numbers. We note that the large numbers of insecu-

rity-induced IDPs in late 2011 and 2012 were coincident with a campaign on the part of the African Union Mission in Somalia forces to oust al-Shabaab from Mogadishu and other areas in southern Somalia. According to some analysts, one of the three major reasons that this campaign was successful was because the populace had withdrawn its support for al-Shabaab in the wake of its perceived mishandling of the previous year's drought (Meleagrou-Hitchens and Solomon 2012). If this reasoning is correct, this example provides a concrete case where the influence of governance during a climate event may have a major impact on US national security interests.

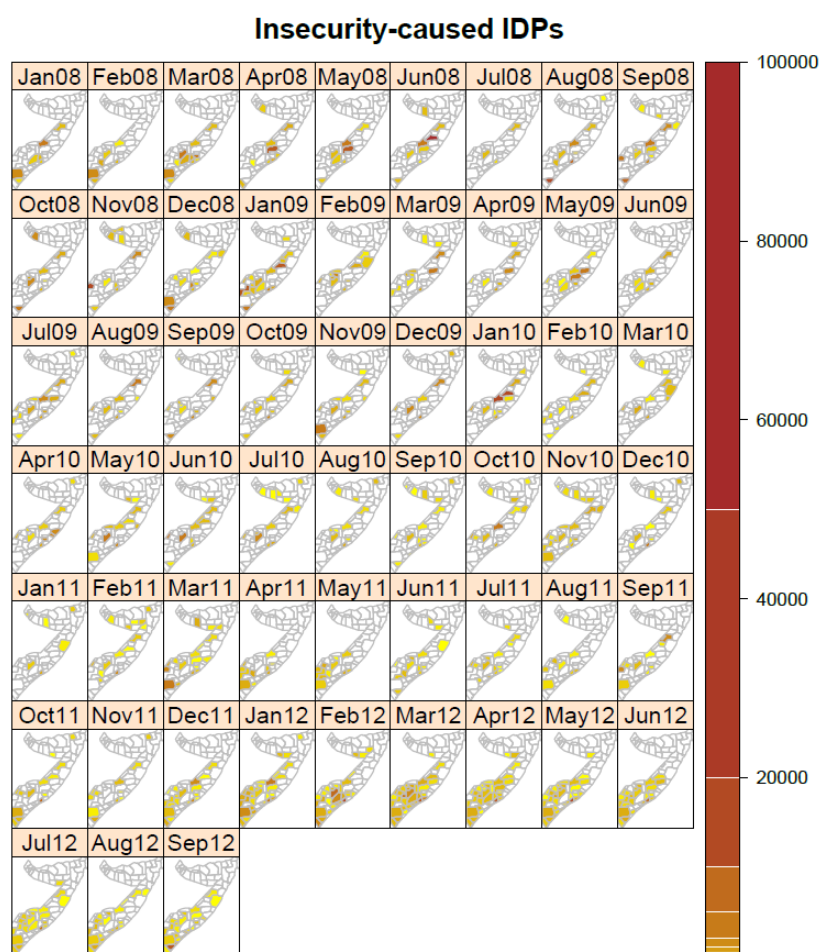


Figure 11. Numbers of insecurity-induced internally displaced persons, by district of out-migration. Note that the scale in this figure is approximately one order of magnitude greater than the scale in Figure 10.

3.3.2 Initial data analysis

An initial comparison of moisture index data and data on drought-induced internal displacements suggests that IDP response to drought as measured by SPI and SSWI is quite variable. As an example of these comparisons, Figures 10 through 15 show the 1-month SPI at a grid cell within Somalia together with drought-induced population movements out of the district over time. In each upper graph, the dashed blue line represents a neutral SPI of zero; and the dashed red line represents an SPI of -1 , below which a drought may be defined to begin (McKee et al. 1993). Each lower graph displays the raw number of people who left the district each month. Figure 12 shows one grid cell in Somalia where the one-month SPI appears to correlate fairly well with population movements, with movements reflecting both the onset and severity of the drought periods in late 2010 and mid-2011. Figure 13, on the other hand, shows drought-induced populations movements that occurred after a lag of several months following a low 1-month SPI in 2008. It also shows only minimal movements following subsequent low SPI months. Figure 14 shows the occurrence of drought-induced population movements despite only minimal drought, as measure by the 1-month SPI. Figure 15 shows another district where significant population movements appear to slightly anticipate, not follow, the 1-month SPI.

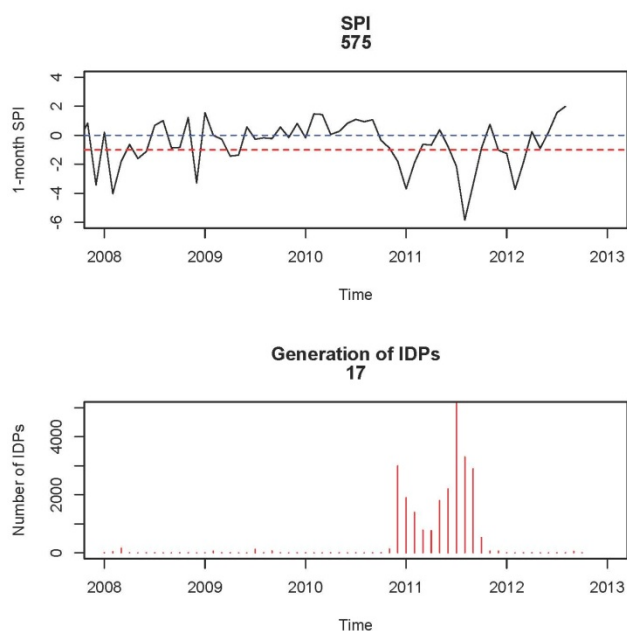


Figure 12. Comparison for one grid cell in Somalia of 1-month SPI and drought-induced IDPs with good correlation.

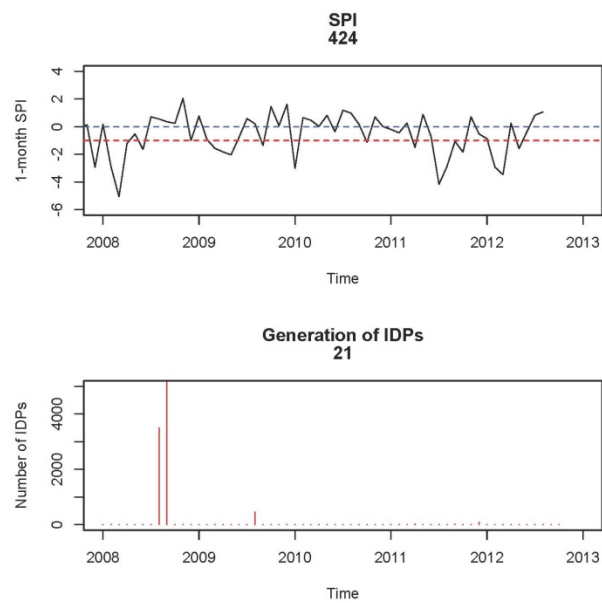


Figure 13. Comparison for one grid cell in Somalia of 1-month SPI and drought-induced IDPs exhibiting an apparent lag in drought response of several months.

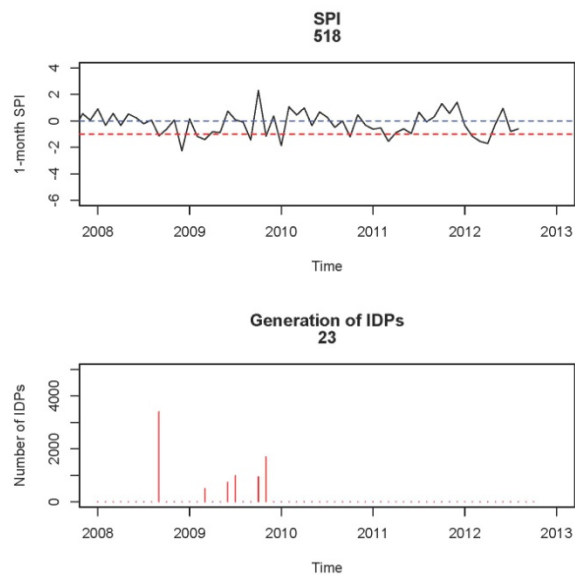


Figure 14. Comparison for one grid cell in Somalia of 1-month SPI and drought-induced IDPs exhibiting apparent IDP response but minimal to no drought.

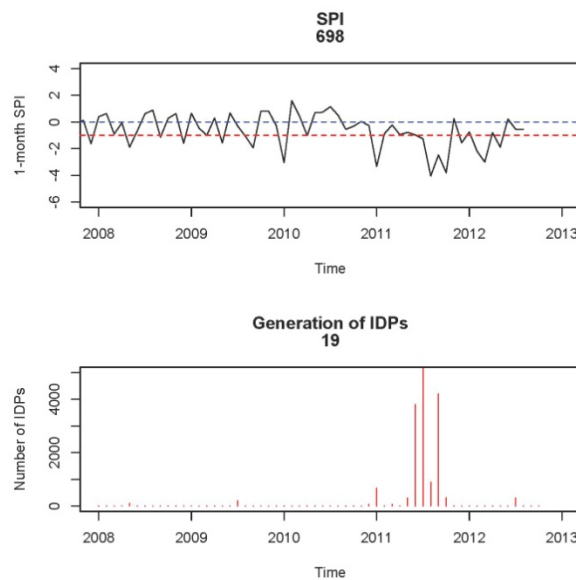


Figure 15. Comparison for one grid cell in Somalia of 1-month SPI and drought-induced IDPs with apparent anticipatory movement.

The data depicted above display information for only one out of 22 calculated moisture indices and 4 out of a possible 1209 grid cells. How do other drought indices in other locations and at other scales correlate with observed population movements? Where is the 1-month SPI highly correlated with movements; and where is another index, perhaps a 6- or 36-month soil water index, for example, better correlated? We attempt to answer these questions by calculating and then mapping correlation coefficients, and we describe the results in Sections 3.3.3 and 3.3.4 that follow.

3.3.3 Correlations between IDPs and moisture indices using Pearson's r

Correlations calculated, with Pearson's r , between moisture indices and drought-induced internal displacements indicate a wide range of correlation values across the set of grid cells covering Somalia (Fig. 16). Each boxplot represents 1209 grid cells; and the large majority of values are negative, indicating that as moisture indices decline, population movements increase. However, median values show that overall correlations are weak. In and of itself, this result suggests that cumulative moisture indices based on the selected retrospective MERRA-Land data at any time scale are not likely to be uniformly good predictors of population movements. As population response to drought is a complex phenomenon that involves more than the mere presence of drought, this result is reasonable and reflects some of the variety in the example input data sets shown in Figures 10 through 15. Notwithstanding overall weak correlations, the strongest

median correlations are associated with the 6- and 8-month SPI and again with the 48-month SPI while the weakest median correlations are associated with the 1- and 18-month SPI. This indicates a certain resilience to the failure of a single rainy month or season but that as drought extends across both the strong and the weak rainy seasons that span a 6- or 8-month time period, increased population movements resulted. It also suggests that relatively long-term (i.e., 48-month) indices are better predictors of movement in this situation than medium-term (i.e., 18- or 24-month) indices.

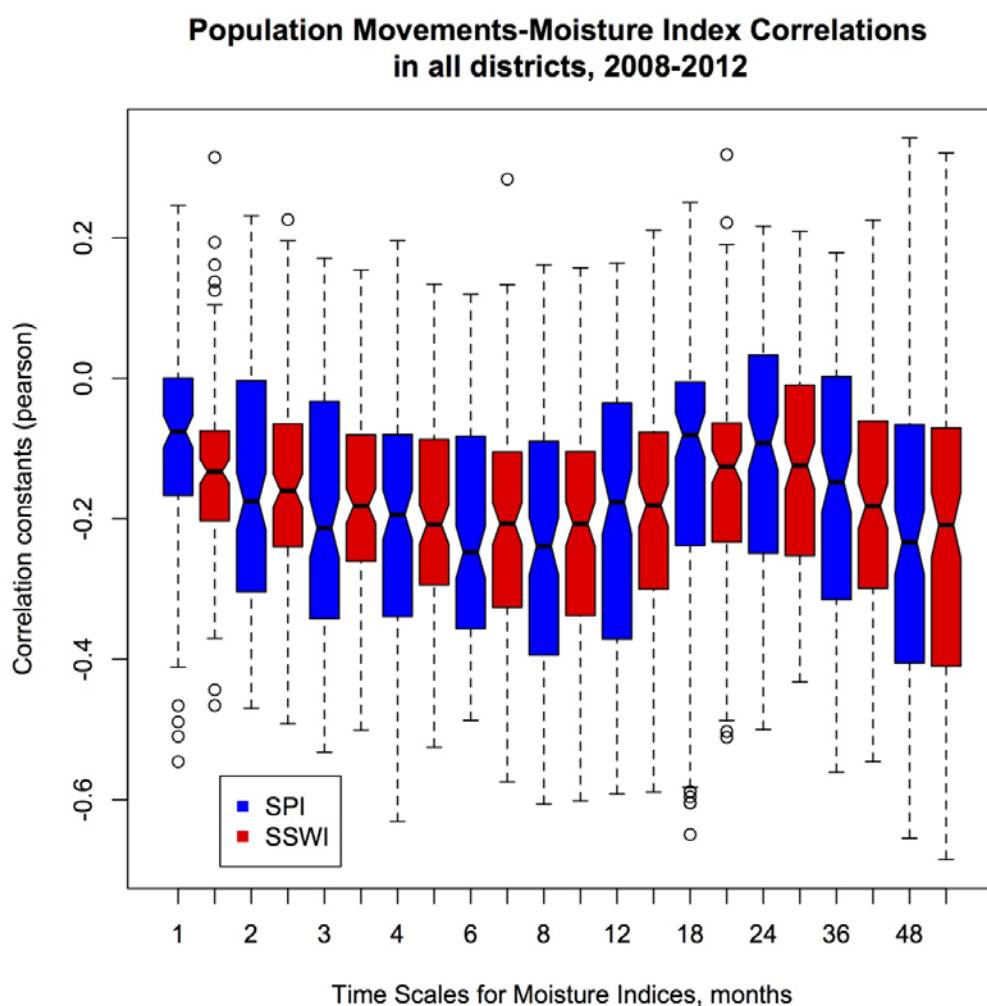


Figure 16. Box-and-whisker diagram showing values of Pearson's r between moisture indices and drought-induced population movements across Somalia between 2008 and 2012. Black horizontal lines indicate median correlation values at each moisture index; boxes indicate 1st and 3rd quartiles; whiskers are drawn at 1.5 times the interquartile range; circles indicate correlation values that fall outside of this range. The hourglass indentations signify the distance of separation required for two median values to be considered significantly different at a 95% confidence level.

However, mapping these correlation values separately for each index shows that the locations at which correlations between moisture indices and population movements are stronger are not randomly distributed. Rather, they are concentrated in the southern and southwestern parts of Somalia, particularly for cumulative indices at time scales of 6–18 months and again for the longer 36- and 48-month time scales (Fig. 17). These results suggest either that (1) the moisture indices are not reflective of the true climate situation in all areas or (2) there are significant geospatial factors of interest that are influencing population response apart from any climate anomalies represented by the moisture indices.

Correlations [pearson] between Drought IDPs and Moisture Indices

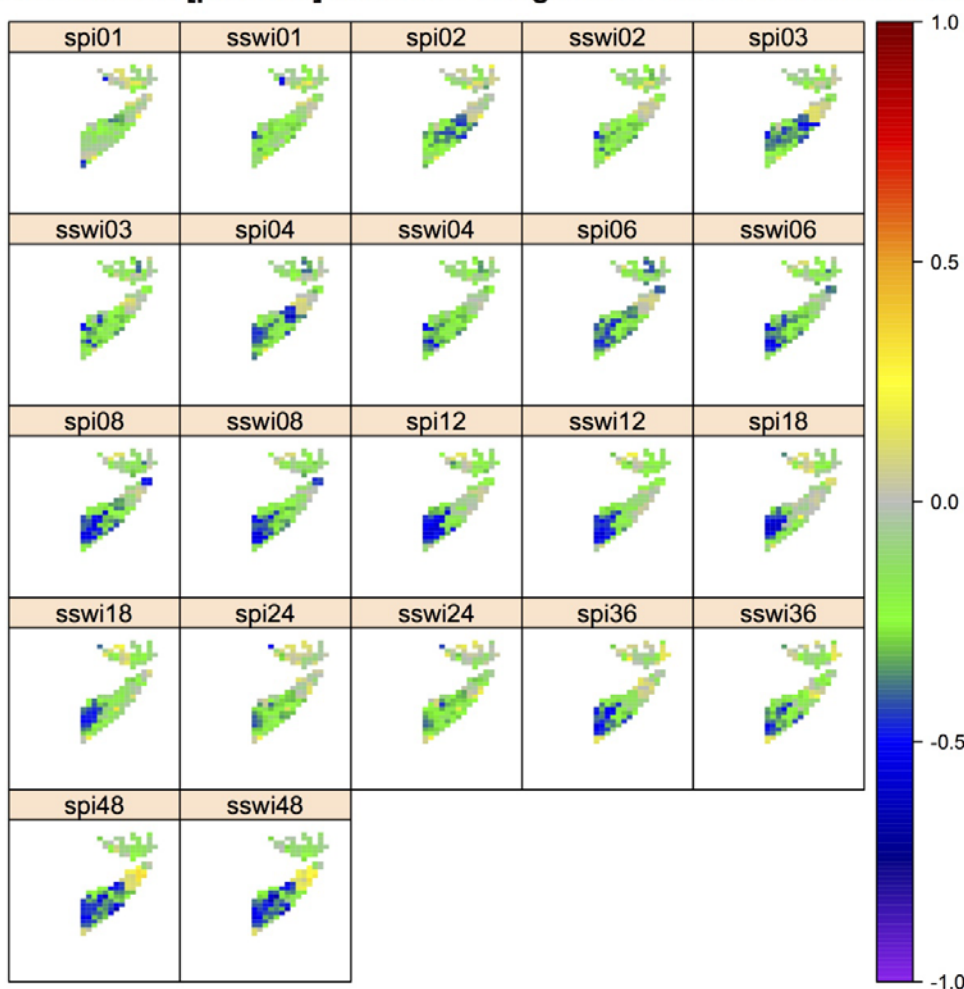
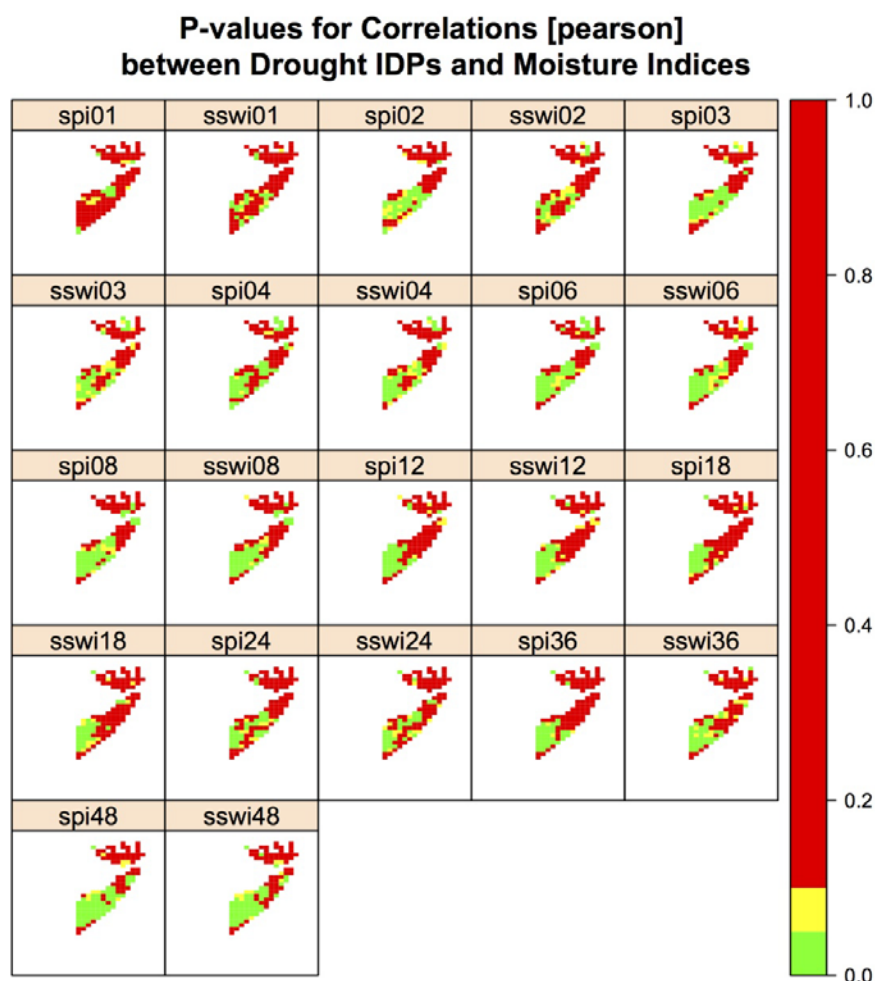


Figure 17. Map of Pearson correlation values between drought-induced IDPs and moisture indices.

As another measure of the effectiveness of the analysis, we calculated p -values to show where the correlations can be considered to be statistically

significant to the 95% level (Fig. 18, green grid cells). These values indicate where, given the number of input data points, results were robust enough that there is only a 5% likelihood or less that the correlation value would have occurred by chance. These results also confirm that, irrespective of the strength of the actual correlation coefficients, results are significant for large areas of non-coastal southern Somalia at a variety of time scales. This graph also helps interpret apparent anomalies in the correlation coefficient map. For example, a 48-month SSWI with positive Pearson r values in central Somalia (shown in yellow in the last panel of Fig. 17) would appear to indicate that as soil moisture increased, population movements due to drought out of the area also increased. An examination of the significance of those values, however, shows that the majority of those cells had values of r that were not statistically significant.



Grid cells in green indicate where correlations
are significant at the 95% level (two-sided)

Figure 18. Significance of mapped Pearson correlation values.

3.3.4 Correlations between IDPs and moisture indices using Kendall's τ

While Pearson's r , because of its tendency to highlight outliers that occur when large numbers of people move, may be a useful statistic for assessing population movement's response to drought, Kendall's τ is a statistically more robust measure for datasets that have a large number of tied values, such as the many months of zero population movement that occur in the PMT dataset. Figure 19 displays a typical distribution of a district's population displacements and is heavily skewed toward these tied values of zero. Because Kendall's τ can compensate for tied values, the same calculations and mapping done with Pearson's r were repeated with Kendall's τ .

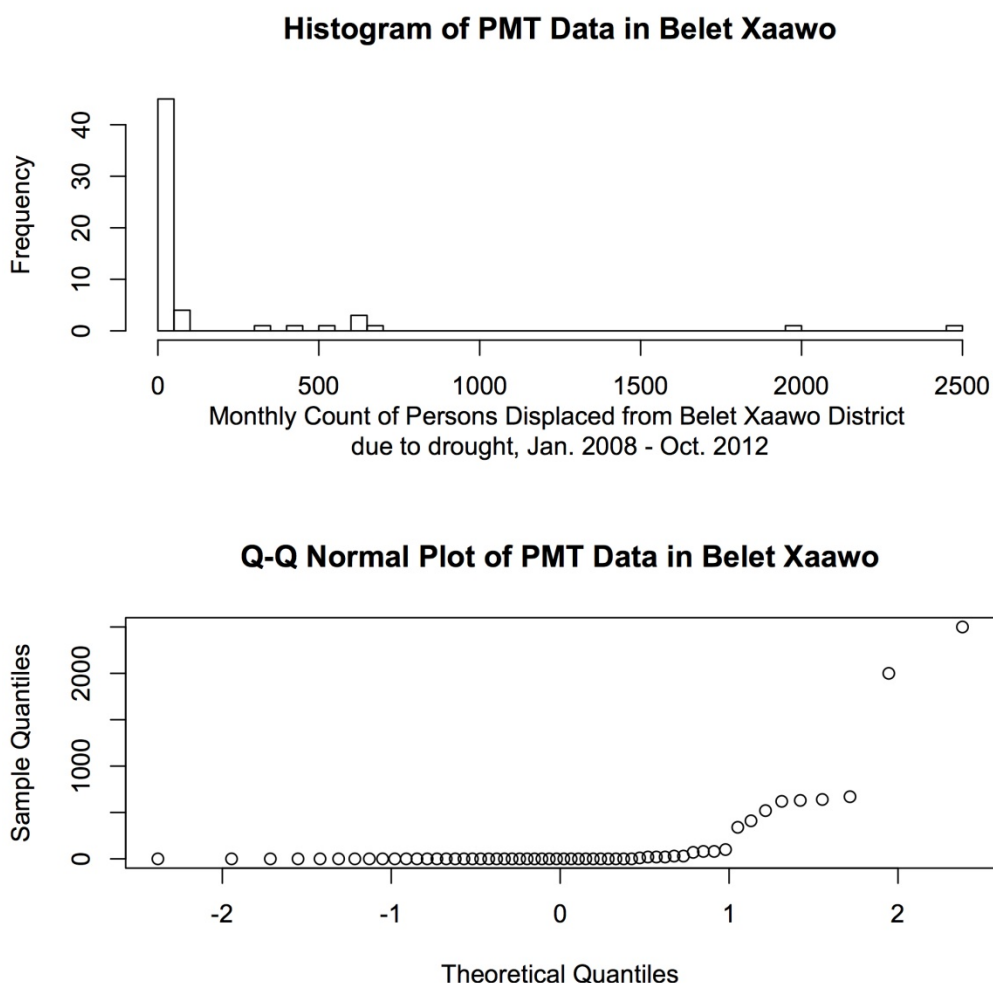


Figure 19. A histogram and quantile-quantile normal plot of IDP movement data in one example district. If normally distributed, the data in the quantile-quantile normal plot would follow a diagonal line through the center of the bottom graph. The graphs show that IDP data is heavily skewed with multiple tied values of zero.

Boxplots showing correlations calculated between moisture indices and drought-induced internal displacements with Kendall's τ , like those of Pearson's r , also indicate a wide range of correlation values across the set of grid cells covering Somalia (Fig. 20). Again, the large majority of values are negative, indicating that as moisture indices decline, population movements increase. It is important to note here that Kendall's τ and Pearson's r values are not strictly equivalent: in general, τ will produce a smaller correlation value than r . For example, data producing a strong r of 0.9 is equivalent to a τ of only 0.7 (Helsel and Hirsch 2002).

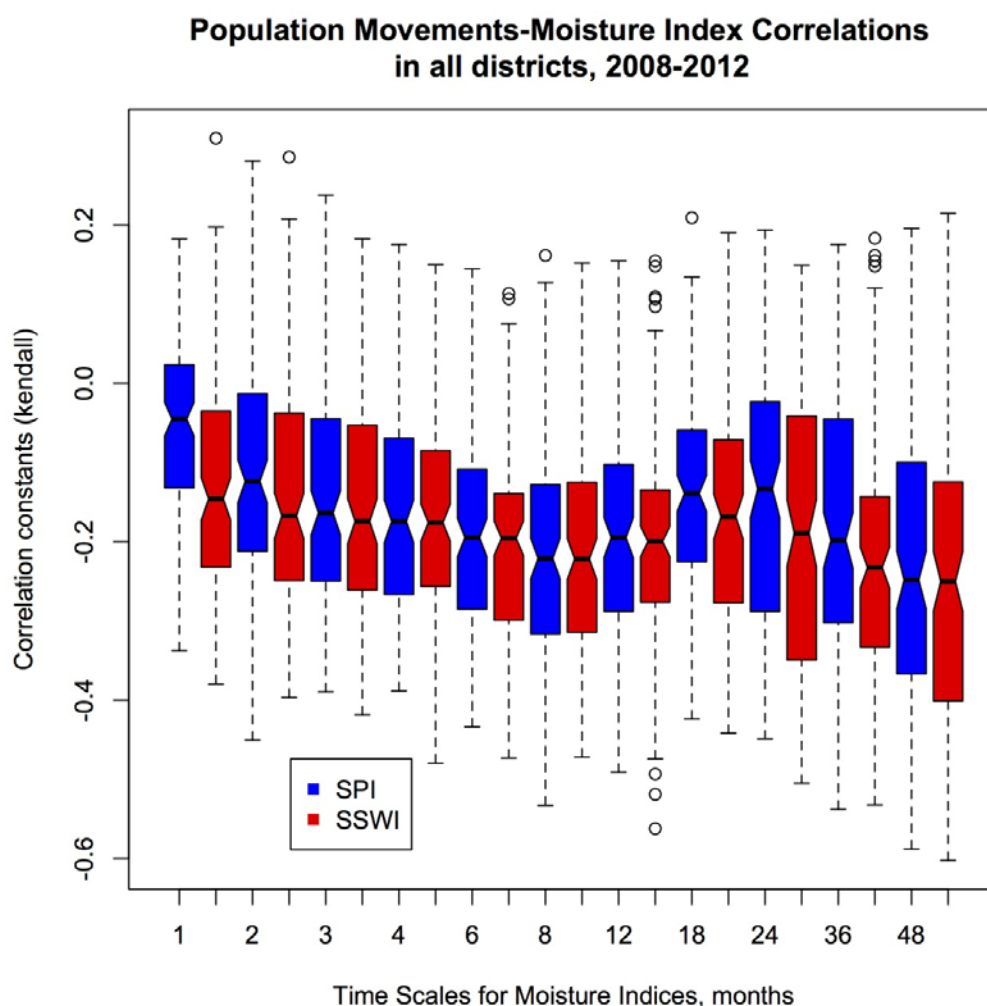


Figure 20. Box-and-whisker diagram showing values of Kendall's τ between moisture indices and drought-induced population movements across Somalia between 2008 and 2012. Median correlation values at each moisture index are indicated with horizontal black line; 1st and 3rd quartiles are indicated by boxes; whiskers are drawn at 1.5 times the interquartile range; circles indicate correlation values that fall outside of this range. The hourglass indentations signify the distance of separation required for two median values to be considered significantly different at a 95% confidence level.

Given this, median values show that overall correlations produced by the Kendall method are somewhat stronger than those produced by Pearson. The Kendall results show that the strongest median correlations are associated with the 8- and 48-month SPI and SSWI, and the weaker values are associated with the 1- and 2-month SPI and with the 18- and 24-month SPI. These results are indicative of decreasing resilience to anomalous dry spells as duration increases up to 8 months. In general, the inter-quartile range is smaller for values of τ than for values of r , indicating less variation across the dataset from grid cell to grid cell. Graphing the locations of each correlation value again shows stronger correlations in the south of Somalia (Fig. 21). The graph of tie-compensated p -values associated with Kendall's τ shows that significant correlations were somewhat more widespread in the south than were the p -values associated with Pearson's r (Fig. 22).

Correlations [kendall] between Drought IDPs and Moisture Indices

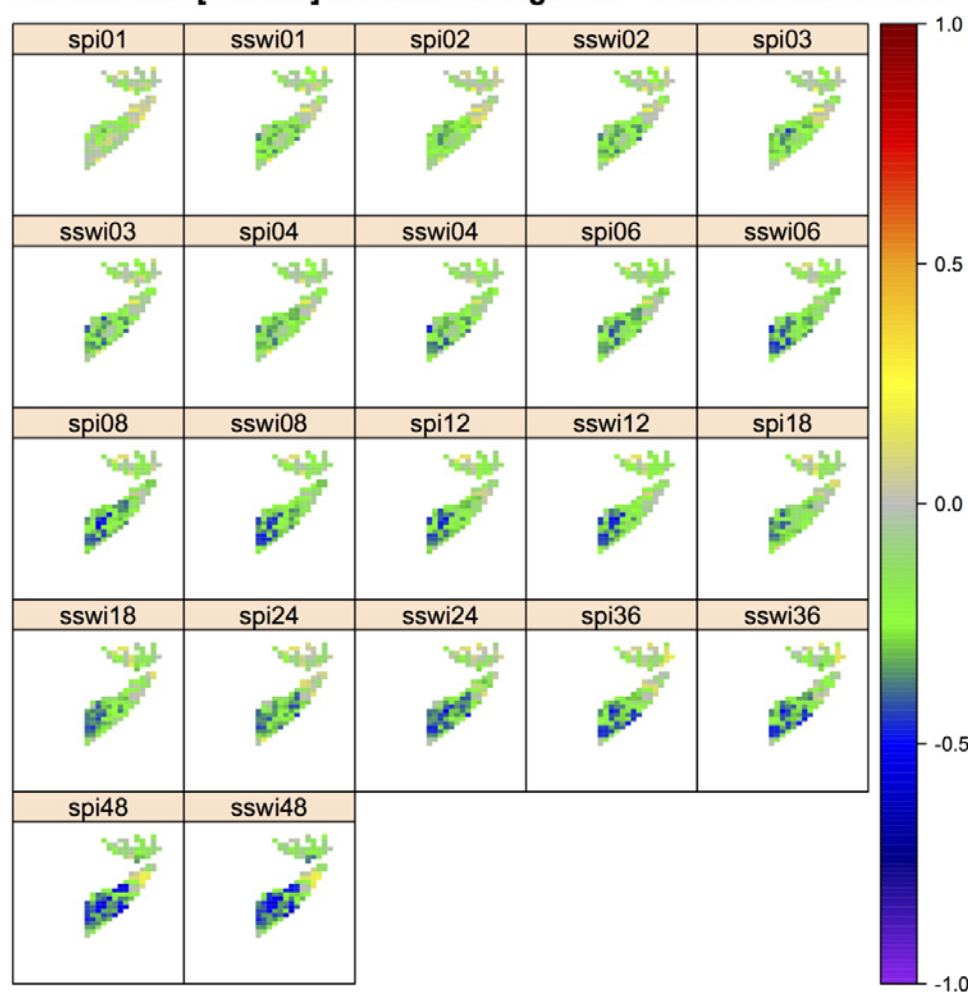


Figure 21. Map of Kendall correlation values between drought-induced IDPs and moisture indices.

Areas where correlations are not significant may also yield useful information. For example, on the shorter time scales of up to 6 months, an area around Mogadishu shows consistently non-significant p -values in the correlation analysis (Fig. 22). This could potentially be an indicator of the de facto extent of the reach of aid and commerce to populations not dependent on subsistence agriculture surrounding the major port city.

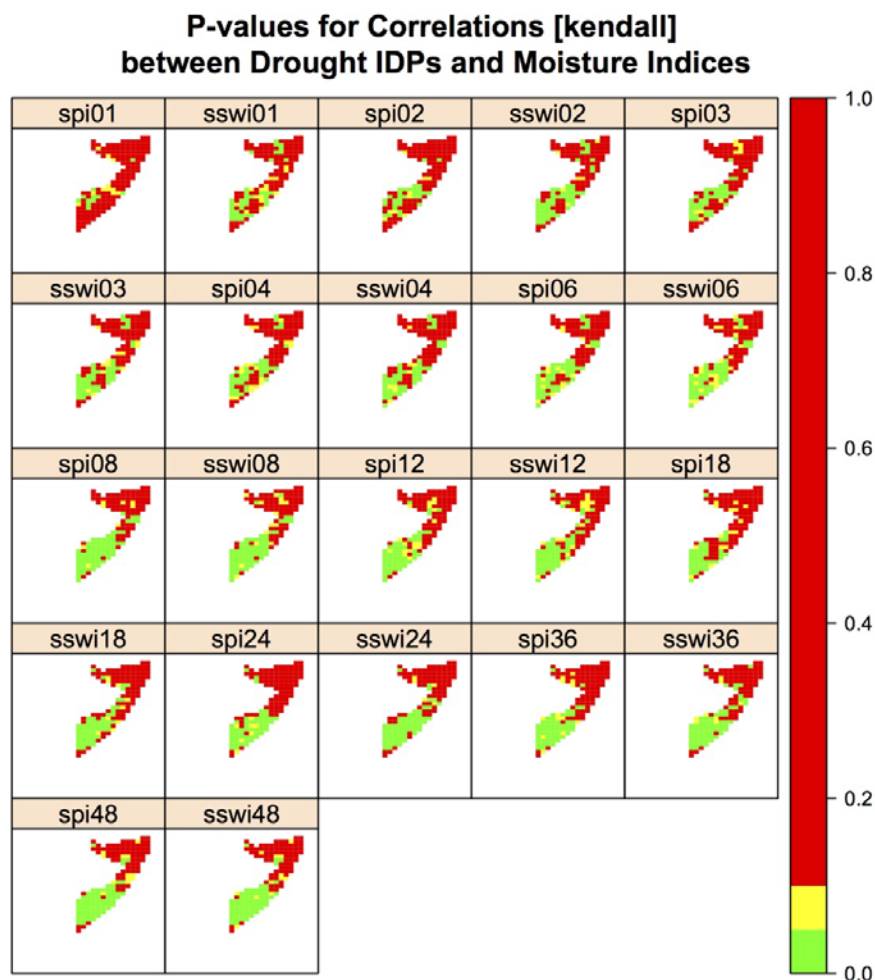


Figure 22. Significance of mapped Kendall correlation values.

3.3.5 Correlations between PMT and moisture indices by livelihood zone

Because of the strong geographic dependence of correlations between moisture indices and population movements, we posited that geographically-differentiated population livelihood mechanisms might partially explain this dependence. Boxplots in Figure 23 show correlation values for sampled points in each livelihood zone. These calculations were repeated

for all moisture indices. The FEWSnet livelihood zone with the strongest (most negative) aggregate correlations between district-level drought-induced population movements and the 8-month SPI was the Bay Agro-Pastoral High Potential Zone, corresponding to livelihood zone 5 in Figure 23. This zone, located in the center of southern Somalia, also maintains a stronger negative correlation than that of other livelihood zones across almost all moisture indices, as shown in Figure 24.

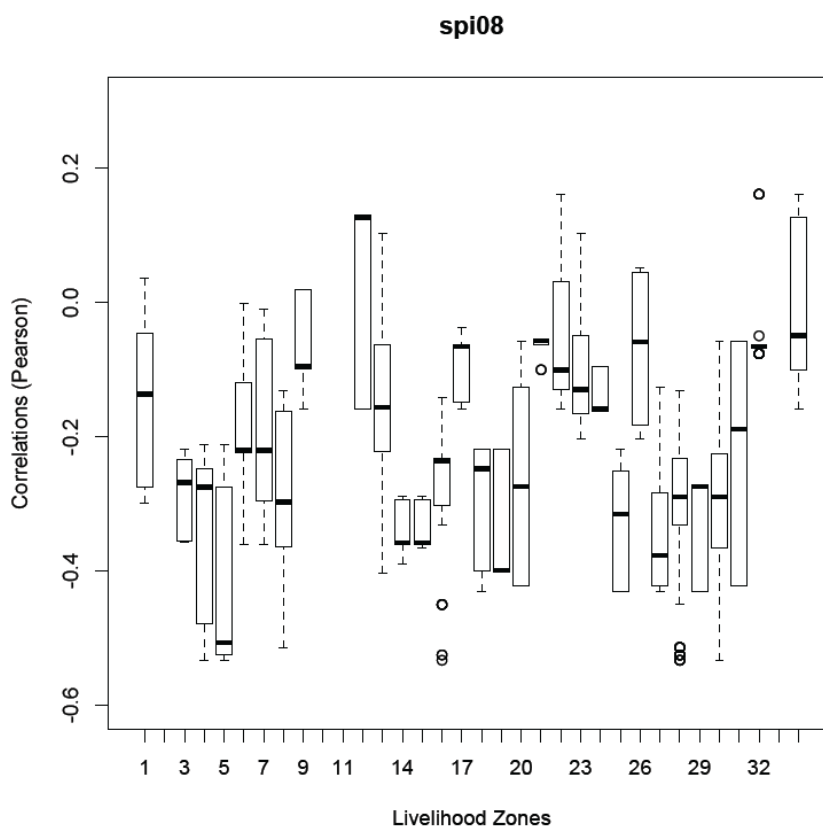


Figure 23. Correlations between population movements and the 8-month SPI, by livelihood zone.

**Correlations between Drought IDPs and Moisture Indices,
by Livelihood Zone**

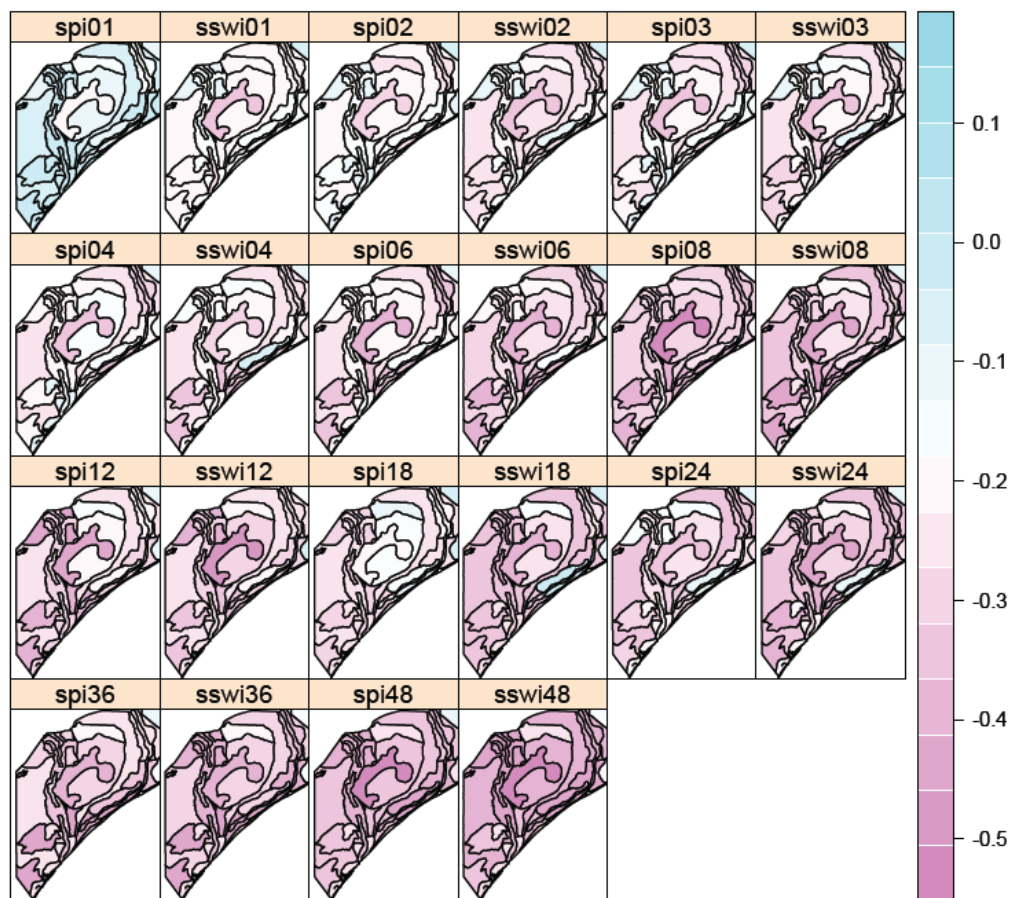


Figure 24. Maps of mean correlations across livelihood zones in southern Somalia.

We note that many of the boxplots in Figure 23 show a highly skewed distribution of correlation values. This is largely due to the disparity of resolution between the MERRA-Land grid cells and the livelihood zones. In particular, riverine and irrigated livelihood zones tend to be elongate in shape or very small and often intersect multiple MERRA-Land grid cells while accounting for only a small fraction of the total area in each. An ideal climate dataset for use in this type of analysis would have higher resolution than MERRA-Land.

Livelihood zones were further aggregated by major type across geographic areas, as shown in Figure 25. From a qualitative perspective, it was not thought likely that this aggregation of livelihood zones would make significant contributions to the previous analysis. The primary principle deter-

minant of population movements appears to be along a geographic north–south division rather than particular livelihood zones (Fig. 26).

Somalia Livelihood Zones, aggregated by type

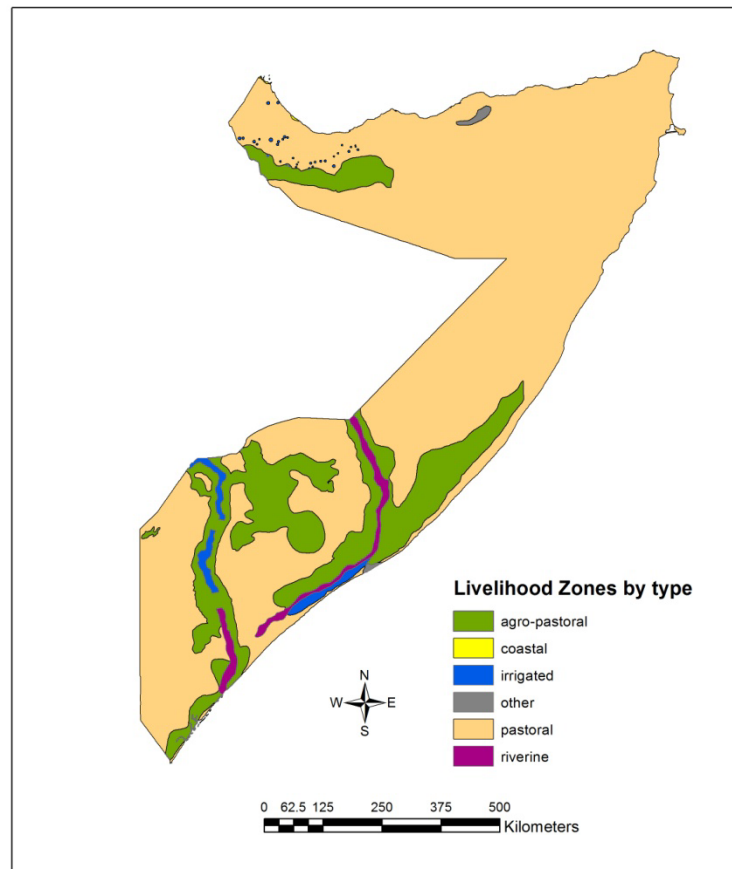


Figure 25. Aggregated livelihood zones by major types.

**Correlations between Drought IDPs and Moisture Indices,
by Livelihood Zone**

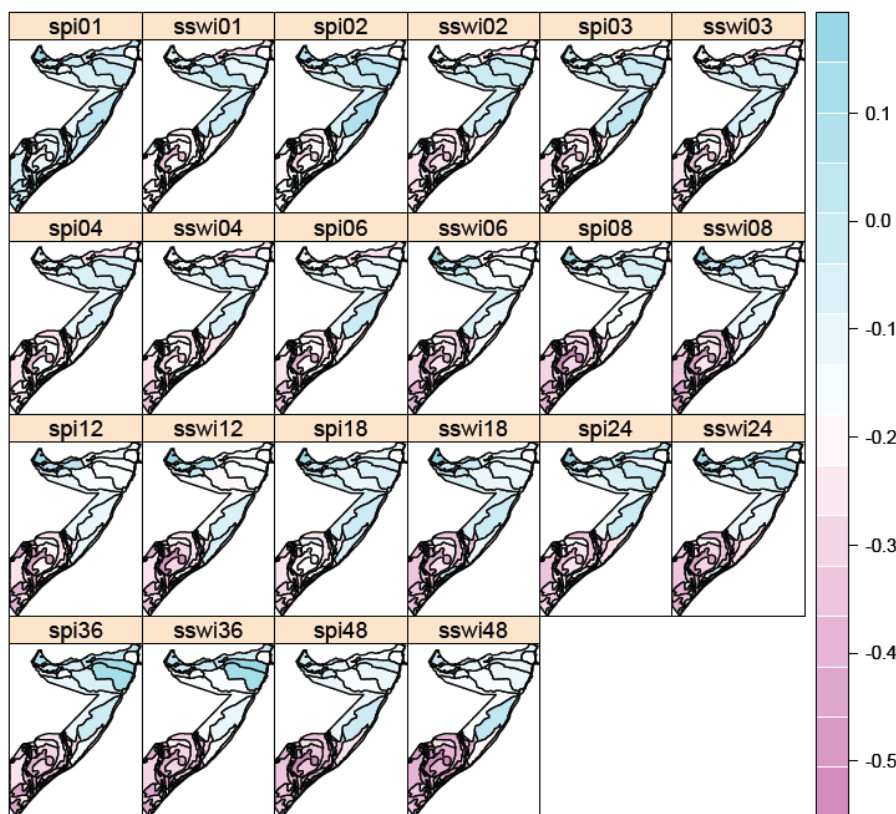


Figure 26. The north–south distinction between correlation strengths does not appear to be related directly to aggregated livelihood zone types depicted in Figure 23.

3.3.6 Linear models

Given certain measurable or possibly predictable drought conditions, what percent of a population might be expected to move due to that drought? This section describes the results of an effort to derive linear models of population movements based on both livelihood-weighted and non-weighted moisture indices in Somalia. An initial comparison of weighted moisture indices from December 2010 to November 2011 based on livelihood-derived critical periods for each district versus drought-induced population movements does not suggest any correlation between the two across Somalia as a whole. An example, given below, of a weighted moisture index based on the 2-month SPI shows a typical data distribution (Fig. 27). Using this index, there appear to be both districts for which movements are close to zero despite low moisture indices during critical periods (as shown by points in the lower left of the graph) and districts with mild to no drought during critical periods that experience significant

population movements (shown by points in the upper right). This suggests that the initial proposed livelihood-based weighting scheme does not adequately distinguish conditions under which populations are subject to moisture stress that leads to migration. Fine-tuning of model weighting parameters could potentially have improved correlations; but over-parameterization without further expert input relevant to the seasonal calendars, however, would not have been justified. Therefore, we attempted to distinguish regions within Somalia that were subject to differing governance and aid regimes.

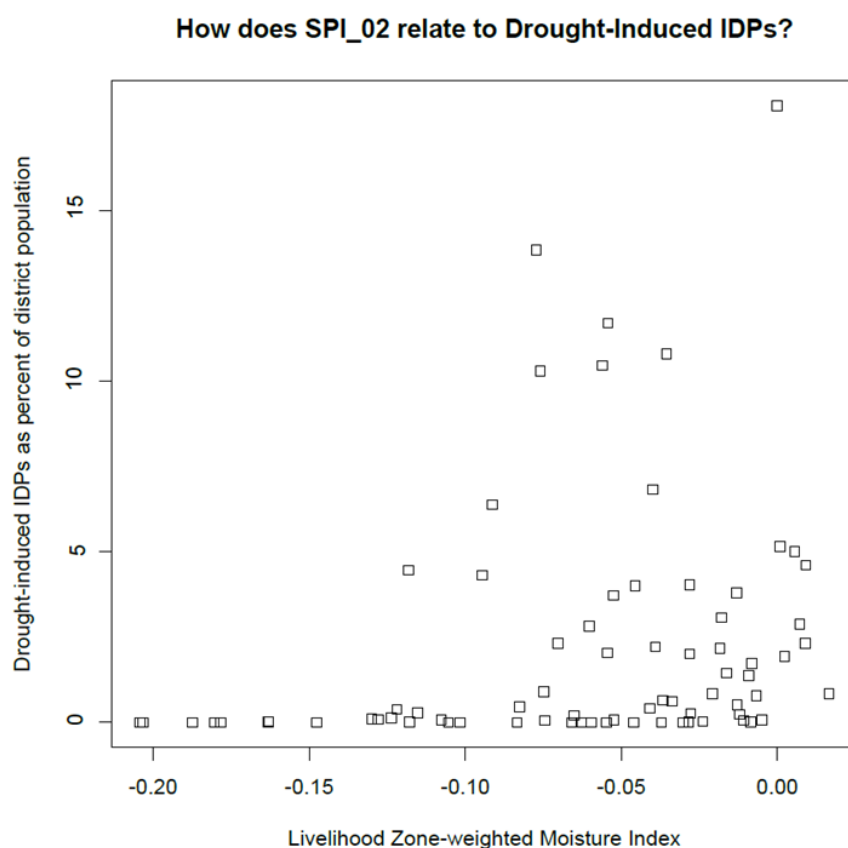


Figure 27. Percent of district population that became IDPs due to drought between December 2010 and November 2011 and corresponding livelihood zone-weighted moisture indices.

3.3.6.1 Influence of al-Shabaab

A set of interrelated influences involving governance in Somalia suggested that the utility of this analysis may improve if we considered separate geographic regions of the country independently. Of particular importance was the radical Islamist group al-Shabaab's control of large areas of south-

ern and central Somalia during the 2011 drought. The World Food Programme (WFP) had pulled operations out of this area by January 2010 due to security concerns for their personnel; and the US government placed al-Shabaab on a list of terrorist organizations, further limiting humanitarian funding and operations to regions under its influence (Hillbruner and Moloney 2012).

To account for this set of factors affecting governance and aid in Somalia during the 2011 drought, we used a map of al-Shabaab influence to define districts within and outside of al-Shabaab influence. In choosing such a map, we established the following criteria: that the map be publically available, dated close to the period of the drought under study, and have a resolution adequate for distinguishing which districts lay predominantly inside areas of al-Shabaab control. It is noted that we made no attempt to verify the accuracy of the particular map used in the current study and that the source of the map is not known with any certainty; the goal was to provide a framework for analysis into which a properly vetted map could be substituted. Given broad similarity in publically available maps, however, the choice of the 2010 map in Figure 28 appears to be a reasonable one.

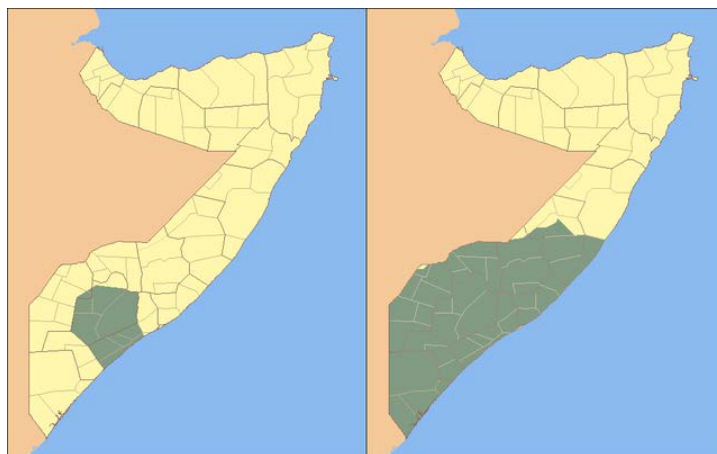


Figure 28. Open-source map of areas of al-Shabaab control in Somalia, December 2009 (left) and January 2010 (right) (Kermanshahi 2010).

A clear distinction arises between population movement behavior inside and outside of areas under 2010 al-Shabaab control, as shown in Figure 29. Here, the livelihood-zone-weighted 2-month SPI moisture index versus drought-induced population-movement data is plotted again, distinguishing districts under al-Shabaab control (red) and those not under its control (black). The two datasets exhibit very different behaviors with population movements occurring almost solely in the southern al-

Shabaab-controlled districts irrespective of drought indices. This result supports the notion that governance mechanisms, whether local or aid-based, play a critical role in the movement response of a population to a major drought. In areas controlled by al-Shabaab, there is evidence at the 90% confidence level that drought-induced internal displacements have a linear relationship to the weighted 2-month moisture index such that the slope of the line is not equal to zero. In areas not controlled by al-Shabaab, there is no evidence that a linear model for drought-induced displacements has a slope other than zero.

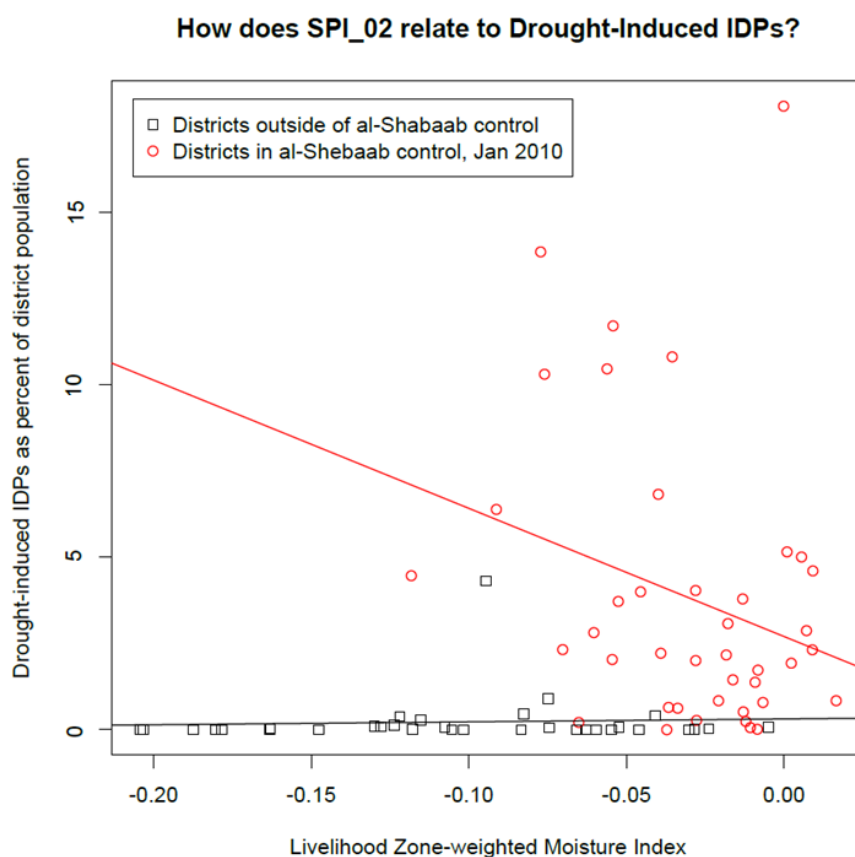


Figure 29. Two linear models for drought-induced IDPs as a function of a weighted moisture index distinguished by governance regime.

This analysis was repeated with both SPI and SSWI at all eleven time scales and for both weighted and unweighted models. The results show that in addition to the 2-month SPI results above, negative slopes significantly different from zero (at the 95% confidence level) were found for 1-, 3-, and 4-month SPI in al-Shabaab-controlled areas for the weighted models. The linear models did not show significance (to greater than 95%) for

any SSWI-based weighted indices or for any areas outside of al-Shabaab influence. The only non-weighted index displaying a statistically significant slope (to the 95% level) of the best-fit line was the 1-month SPI. Table 3 summarizes these results.

Table 3. Summary of significance of slope values for different linear models. Green and yellow cells indicate which indices are associated with a slope significantly different from zero (to the 95% and 90% level, respectively). White cells indicate that there is no significant relationship in this analysis for that moisture index. The linear models could be used to estimate potential drought-induced IDPs as a function of a weighted moisture index.

Type of moisture index	Applied to areas inside or outside of al-Shabaab control?	Livelihood-Zone-Weighted index	Non-weighted index
1-month SPI	Inside	95%	95%
2-month SPI	Inside	90%	<90%
3-month SPI	Inside	95%	<90%
4-month SPI	Inside	95%	<90%
5-month SPI or longer	Inside	<90%	<90%
Any SSWI	Inside	<90%	<90%
Any index in an area outside of al-Shabaab control	Outside	<90%	<90%

While Sections 3.3.3 and 3.3.4 indicated that the shorter-term indices were not as well-correlated with population movements as some of the longer-term indices, this analysis suggests that use of the shorter-term indices to analyze IDP behavior over the time period of one year might still yield some useful information. The fact that one version of the non-weighted index also produces high confidence levels suggests that the livelihood-zone distinction is not essential. We note that the use of the seasonal calendars in the livelihood-zone weighted analysis inherently favors the use of SPI rather than SSWI as the seasonal calendars demarcate rainy seasons and not seasons of favorable soil moisture, which would tend to continue beyond the end of a rainy season.

A separate study on seasonality patterns in Africa (Herrmann and Mohr 2011), however, suggests that the association between al-Shabaab-controlled areas and drought-induced migration may not point to al-Shabaab and governance being the only explanatory variable for the spa-

tial distinction in population response to drought. Figure 30 shows a seasonality map of Africa, classified by number and modality of wet seasons, derived from remotely-sensed rainfall and temperature data. The areas of southern Somalia that typically experience two unimodal wet seasons are quite similar to the areas that this study found to have had significant drought-PMT correlations and were also similar to the areas under al-Shabaab control in 2010. This complicates the reasoning that governance factors are the major explanatory variable distinguishing the two linear models above. The differences in seasonality classes could indicate the following:

1. Seasonal moisture anomalies used in this study are ineffective for understanding population movement response to drought in arid regions, and a drought indicator that took seasonality classes or absolute moisture values into account may provide additional useful distinctions in this type of analysis.
2. There may be underlying social and economic factors that limited the extent of al-Shabaab influence to the regions of southern Somalia in which two unimodal wet seasons are the norm, suggesting the possibility of a more complex historical or anthropological relationship between climate and the evolution of governance structures.
3. The spatial similarity of these governance and seasonality datasets could be purely coincidental.

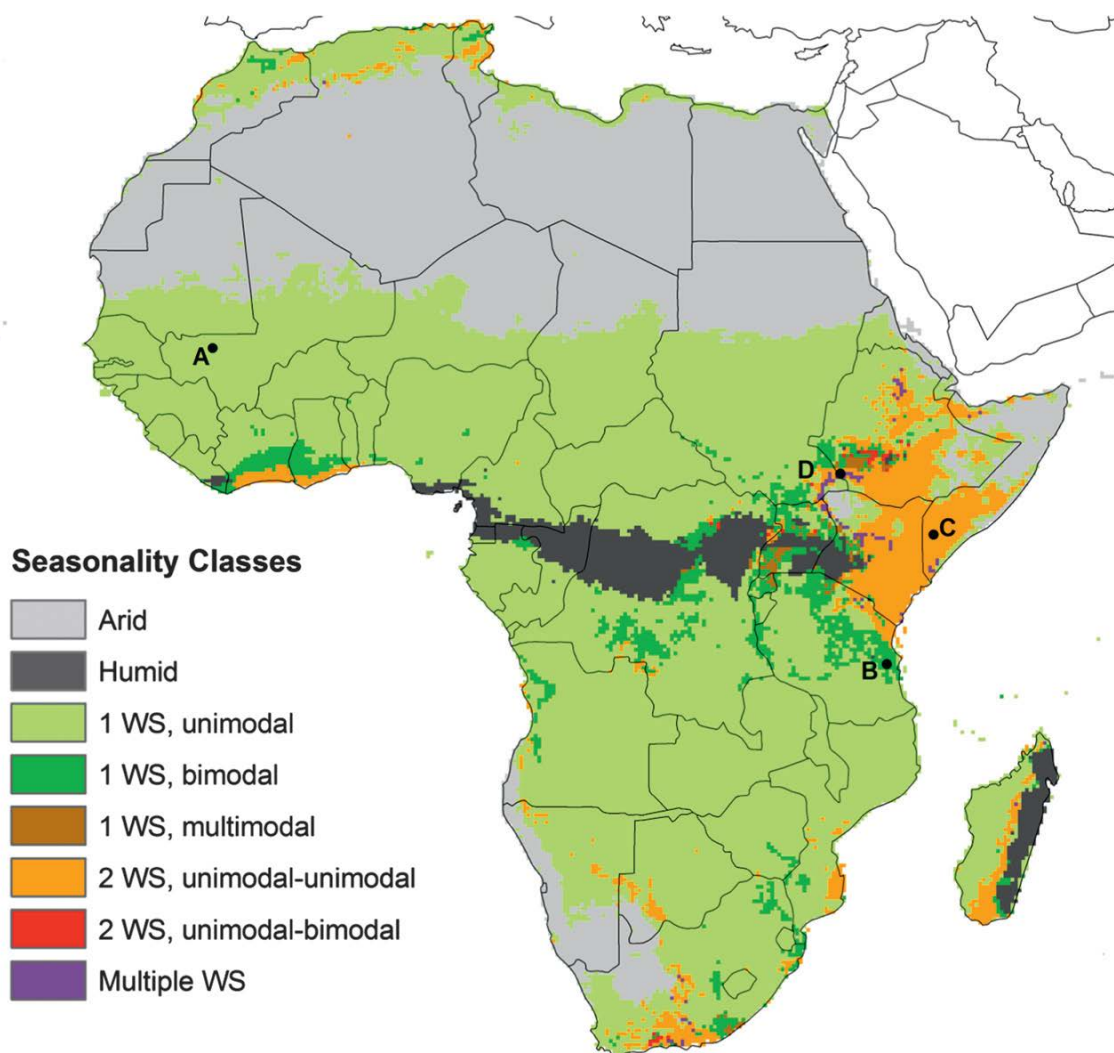


Figure 30. Seasonality map of Africa classified by number and modality of wet seasons and derived from remotely-sensed rainfall and temperature data (Herrmann and Mohr, 2011).

3.3.6.2 Linear models in predictive mode

Subsequent analysis returns to the linear models developed for southern Somalia. Given a linear model and an input dataset, the model can then be run in “predictive” mode, displaying projected confidence intervals surrounding the modeled values of fractional population displacement due to drought. As an example, Figure 31 displays the 90% confidence intervals around the weighted moisture index based on the 2-month SPI for the PMT dataset from December 2010 to November 2011 in al-Shabaab-controlled areas. This projection suggests that during a year with a mean weighted 2-month SPI of -0.04 , a district could lose approximately 3.5%–5.5% of its population. However, the same projection suggests that even when the mean weighted moisture index is zero and no drought is present,

we can expect some population movement. As this weighted drought index intensifies, the percentages of displaced population increase to around 8%, but the uncertainty surrounding that estimate is substantial.

Because we derived this model from a specific set of observations, we could use it only to “predict” population displacements in situations that are adequately similar to the one from which the data was derived. In reality, this may never be the case due to the complex situation in this location. However, this analysis nonetheless represents an empirical effort to quantify the effects of a specific climate-related stress on population movements, taking into account mechanisms of both livelihood and governance.

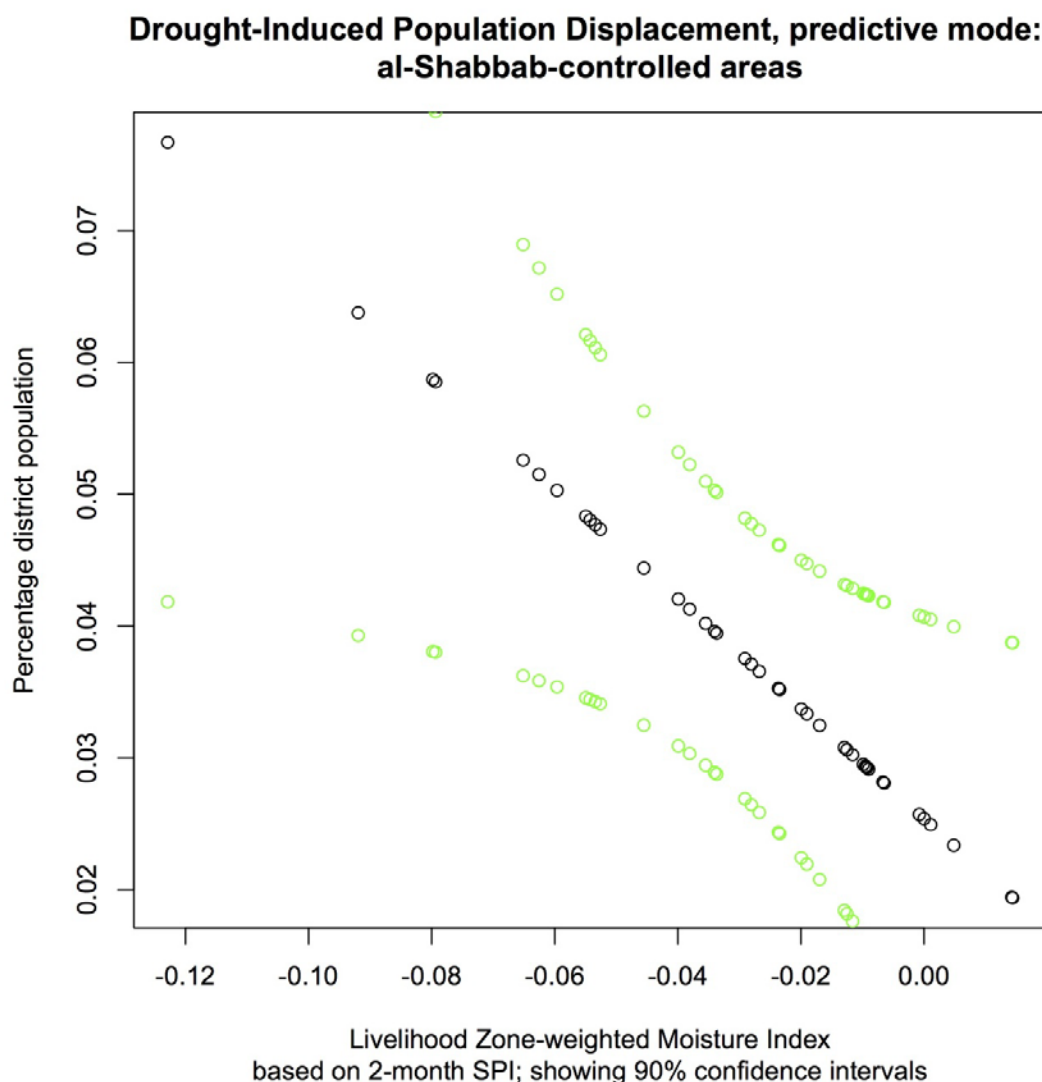


Figure 31. The 2-month SPI weighted model was run in predictive mode to depict confidence intervals.

3.3.6.3 *Outlier identification and residual mapping*

We used several techniques to identify districts that may be considered outliers in these linear models (Fig. 32). A plot of residuals versus fitted values should show residuals to be randomly distributed across all fitted values both above and below a horizontal line at a residual of zero. Similarly, a scale-location plot should not show any obvious trends. A quantile-quantile plot should show a normal distribution of residuals with data points falling on the 1:1 diagonal. Finally, plots of residuals versus leverage should highlight districts that have unusual amounts of leverage on the results of the model. In this case, and more generally for all other cases examined, no data points were found outside of a Cook's distance of 1 (Chatterjee and Hadi 1986), indicating that no districts that might be outliers have an undue amount of influence over the linear model parameters. Some of the data points that lie close to this distance in the models, however, are associated with districts with small population counts or are near important international borders. In these cases, low population counts in the LandScan data set could magnify the displaced population expressed as a fraction of district population. In addition, one district near the Ethiopian border was classified as being in al-Shabaab-controlled territory, but was not entirely so according to the Kermanshahi map, and likely functioned as a safe haven that retained access to international aid rather than a district from which people were forced to leave during drought. Figure 33 show an example map of model residuals in southern Somalia for the 1-month SPI.

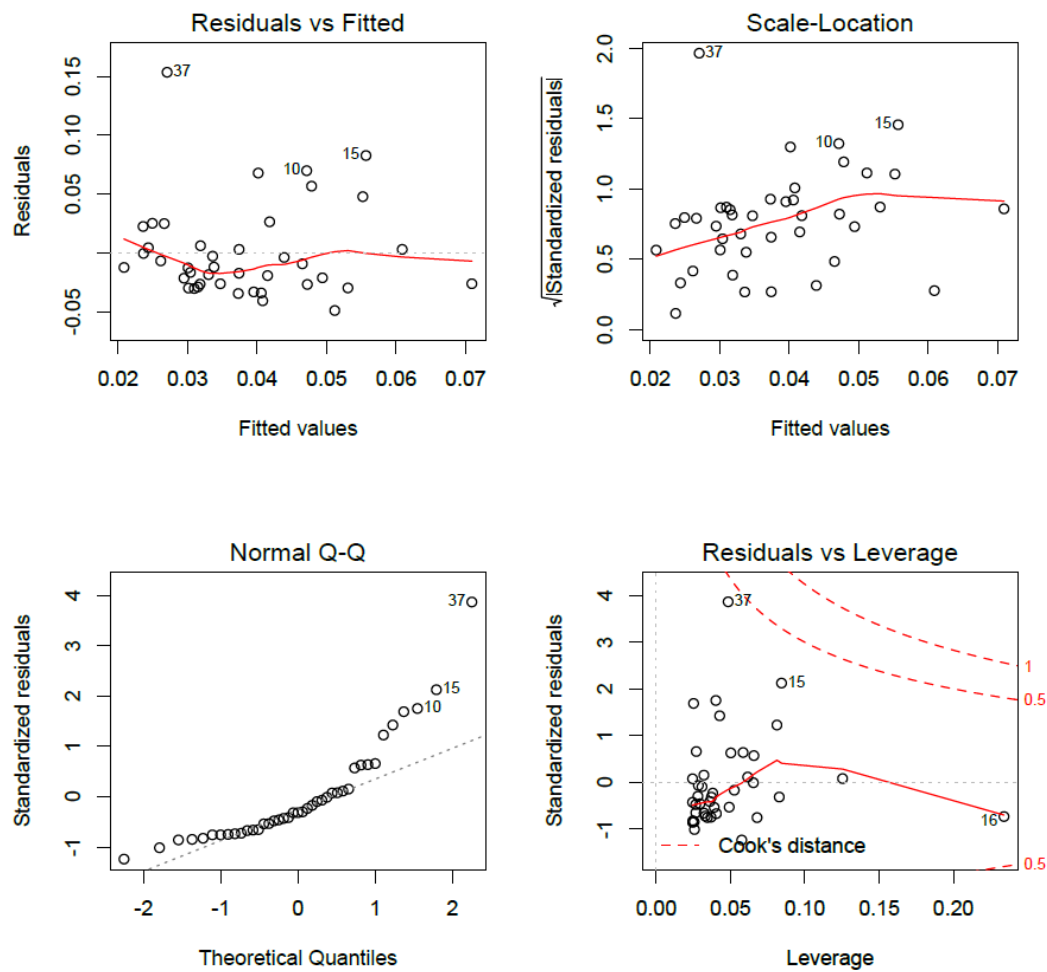


Figure 32. Four methods for assessing outliers in a linear model.

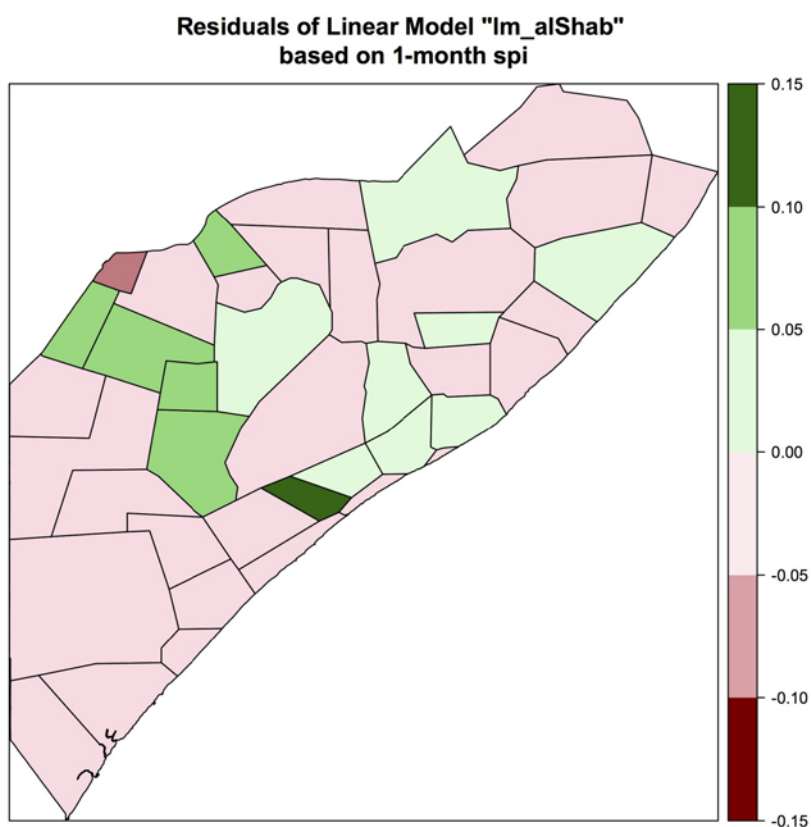


Figure 33. Residuals map for the livelihood-zone-weighted linear model of drought-induced population response to mean 1-month SPI in al-Shabaab-controlled areas. The scale bar depicts residuals as fraction of total district population. The districts with the highest and lowest residuals were, respectively, a district with a low population where percent population displacement was underestimated by the model (dark green) and a district on the Ethiopian border that was not entirely under al-Shabaab control and where population displacement was overestimated by the model (dark pink).

We found that removing some of the outliers identified by these methods improved the p -values of the linear models run on the data sets. However, without additional expert input or a well-defined method to determine which districts should be tagged as outliers and removed from the analysis, removal would subtract from the overall simplicity of the method and would lead to an over-parameterized model. The linear models developed here within the al-Shabaab-controlled districts provide a general estimate, but not an accurate spatially-disaggregated picture, of population movements in response to drought in this area.

3.4 Market price network analysis

Market price data sets could potentially be used for a variety of purposes. FEWSnet and FSNAU use this data primarily to understand price patterns

for early warning of famine and for other humanitarian assessments, and FSNAU produces detailed regional assessments twice a year following the rainy seasons (FSNAU-Somalia 2013). The purpose of our analysis is not to repeat this type of assessment but rather to determine whether this data might also contain information of economic interest related to the provision of security. For example, which markets are particularly well connected and might, therefore, be of particular importance to maintain safe corridors between? Which markets are not well connected and might, therefore, be particularly vulnerable to population movements during local drought conditions? Does market connectivity change significantly from drought to non-drought years? How tightly correlated are prices to drought severity for different products and locations, and what kind of lag times are typical between onset of drought and price changes? Are there anomalous price behaviors in some markets that might indicate the presence of other economic influences, such as influx of aid or presence of illicit activities? In our study, we use several statistical and network techniques in attempts to provide answers to some of these questions.

3.4.1 Market price trends

Of the 33 products whose prices were monitored by FSNAU at 44 market locations, several groups of products appeared to have price fluctuations related to the onset or deepening of the 2010–2011 drought. Cereals and legumes, such as red and white sorghum, wheat, white maize, and cowpeas, exhibited price increases during this time period while imported red rice remained somewhat more stable (Fig. 32). Livestock prices, such as export-quality sheep, cattle, and goats, exhibited an initial drop in prices in certain markets toward the end of 2010 when the drought began and a general rise through the end of 2011 as it ended (Fig. 35). Diesel and petrol prices rose in mid-2011 while the day labor rate remained on average somewhat more stable (Fig. 36).

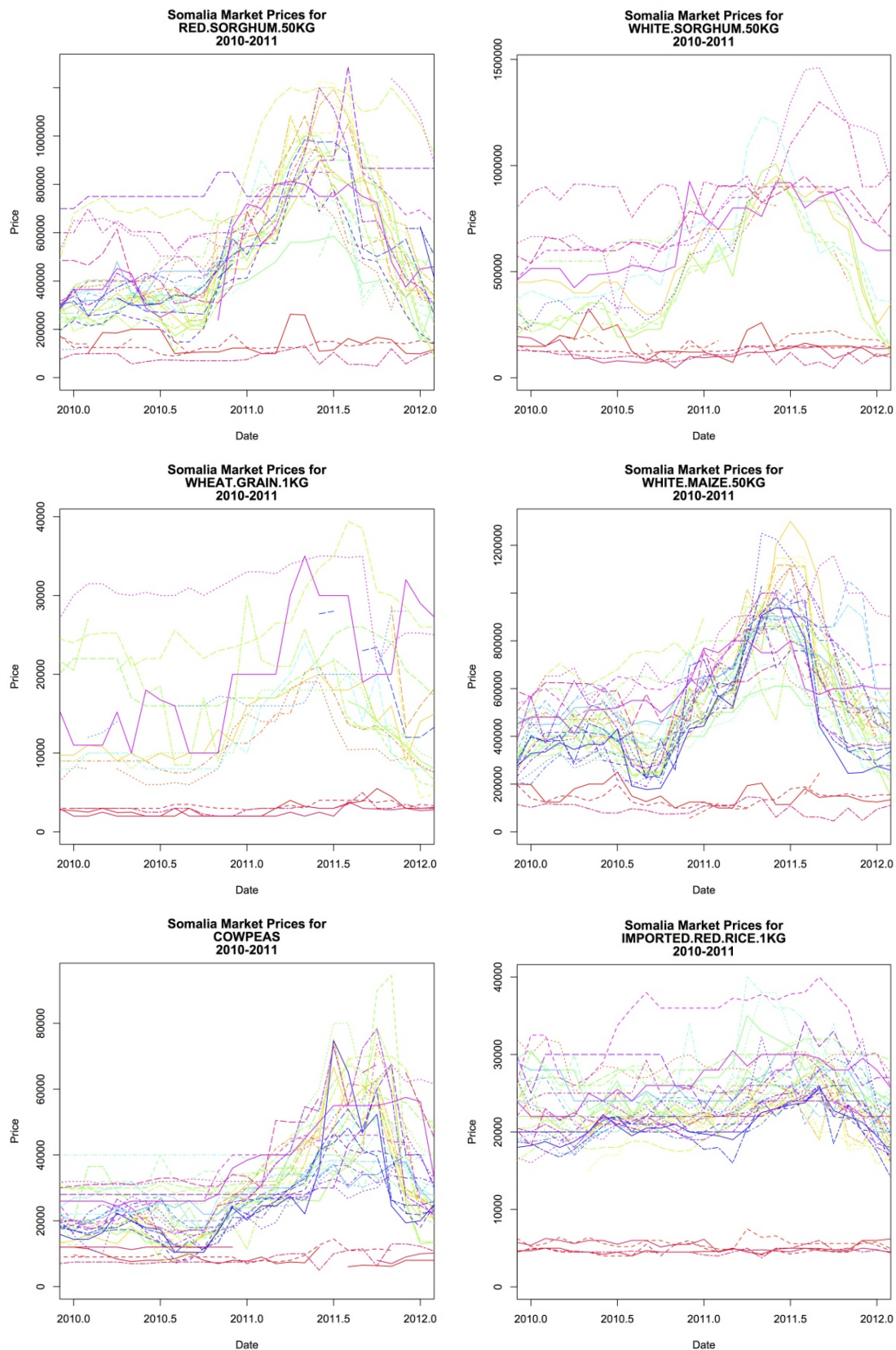


Figure 34. Prices for cereals and legumes, 2010–2011 Monetary denominations vary.

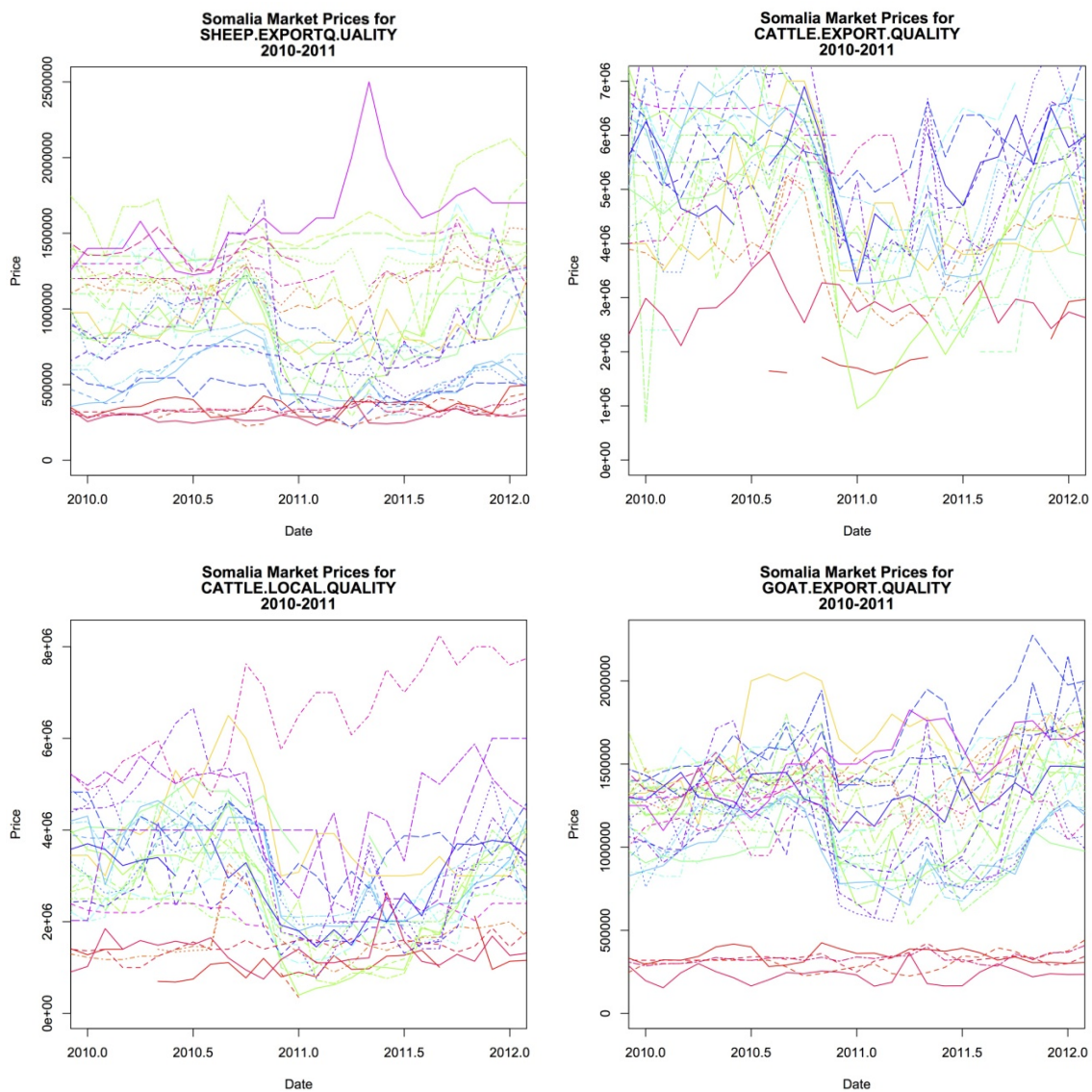


Figure 35. Market prices for livestock, 2010–2011. Monetary denominations vary.

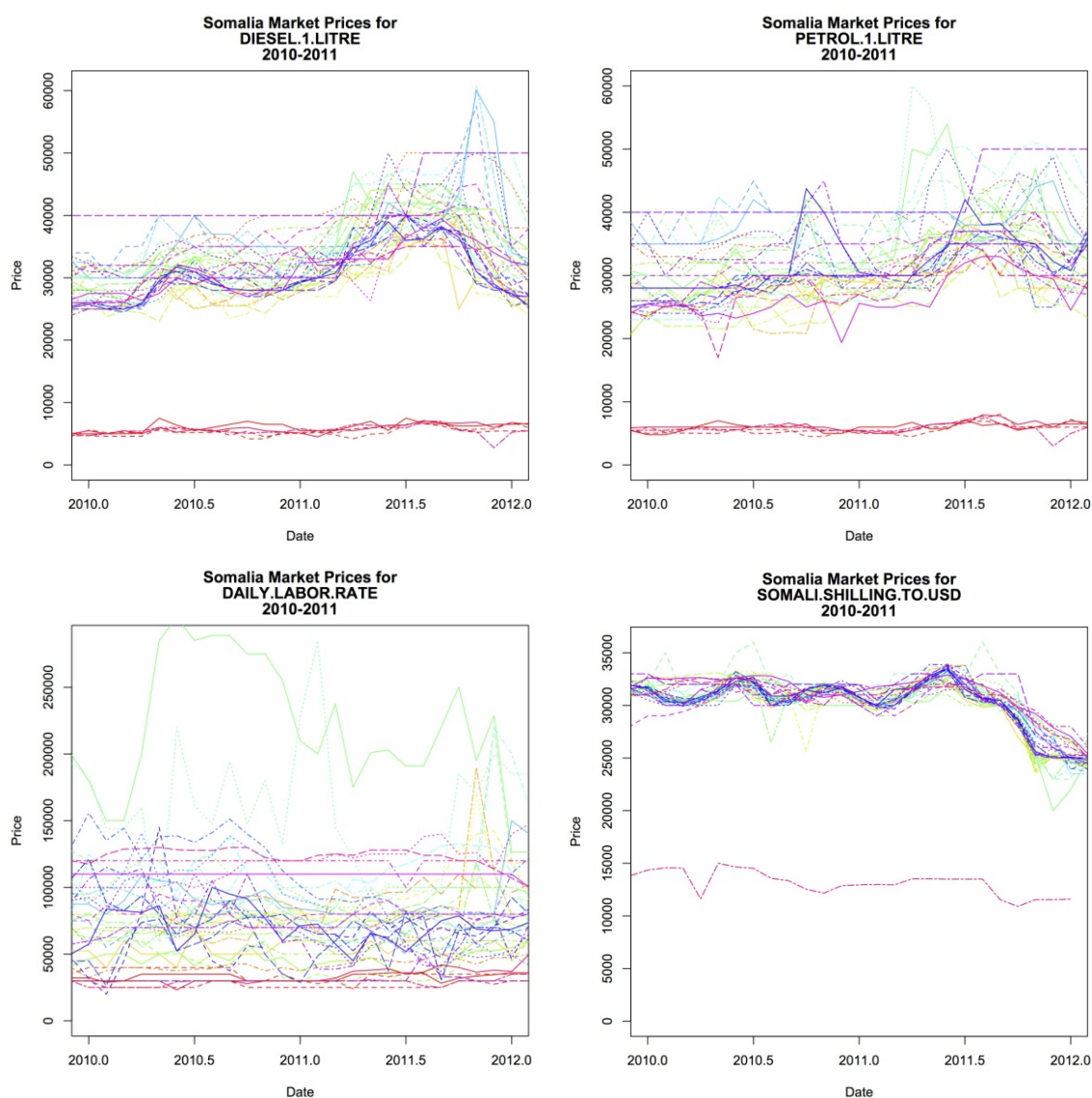


Figure 36. Market prices for fuel and the day labor rate and the exchange rate for the US dollar and the Somali Shilling, 2010–2011. Monetary denominations vary.

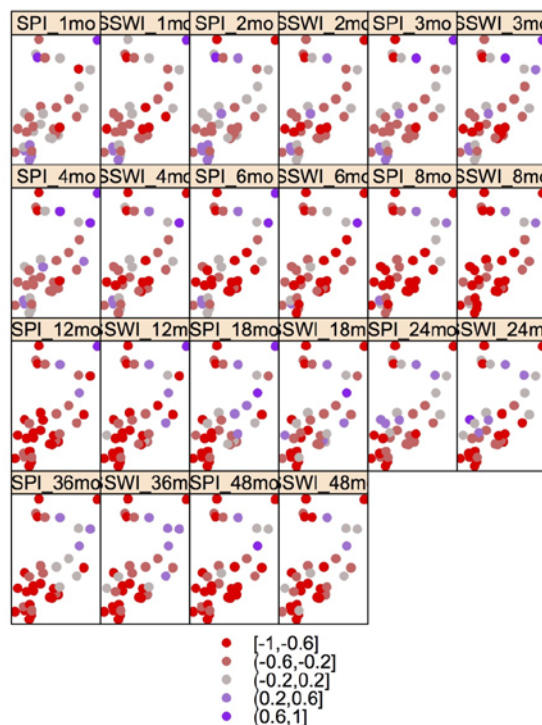
3.4.2 Correlations between prices and moisture indices

Drought is a major cause of price fluctuations in locally produced cereals and livestock, and understanding which moisture indices tend to be most associated with price fluctuations in a given region could be useful for anticipating a wide variety of population responses to drought. While agricultural economists use crop and livestock models that use crop growth and weather inputs to understand price cycles and fluctuations, the simple correlations used in our study attempt to identify overall effects without the more detailed biological inputs used in agricultural models.

Our correlation analysis, described in Section 2.4.2, found that the strongest relationships between moisture indices and price-to-day-labor ratios occurred for two staples, white maize and red sorghum in southern Somalia during the December 2010–November 2011 period (Fig. 37). The strongest negative correlations peak near 8-month indices and again near longer-duration indices such as the 48-month SPI. These price ratio increases during drought for these two staple products appear to be broadly consistent with patterns seen in previous drought-population movement correlation data (e.g., Fig. 21 and 22).

These results were not uniform across all products, however; and we note several points that qualify the utility of this analysis. First, the availability of the price data differed from product to product, so it was not possible to look at all products during the same time periods. Second, a simple correlation clearly does not consider any external economic factors, such as livestock import bans or regional fuel prices, that may determine prices and price ratios. Third, using the moisture indices only at the location of a market in this analysis presumes that the entire market catchment from which products are derived has experienced similar climate conditions to the market location itself. This makes the analysis most appropriate for those products produced locally and not widely exported, such as staple products of subsistence agriculture.

**Correlation Strength
between WHITE.MAIZE.50KG price/labor ratios and Moisture Indices**



**Correlation Strength
between RED.SORGHUM.50KG price/labor ratios and Moisture Indices**

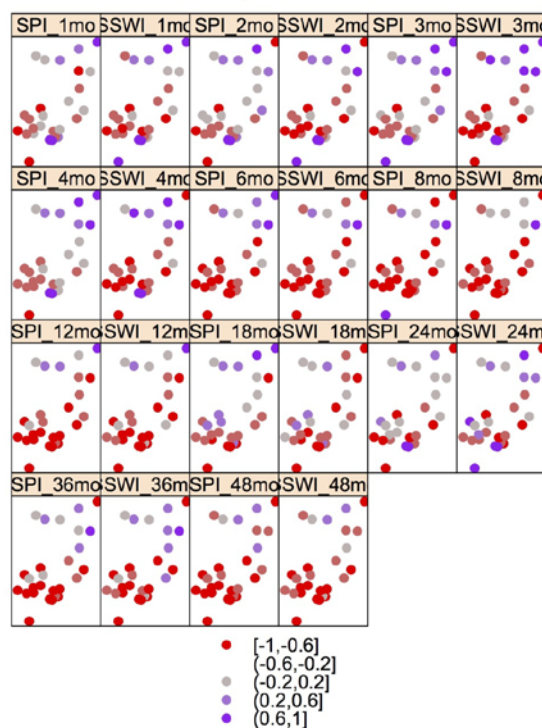


Figure 37. Southern Somalia in particular exhibits strong negative correlations indicating price-ratio increases with drought between December 2010 and November 2011.

3.4.3 Lag times for market prices and drought

The CCACF analysis produced maps of statistically significant lag times between the price-to-day-labor-rate ratio for each market product and 1 of 16 different moisture indices (SPI and SSWI for 1, 2, 3, 4, 6, and 8-month periods each). The analysis was run at a variety of time scales, from the full duration of the FSNAU market price dataset (1995–2012) to certain year-long subsets of that data. A complete presentation of all results is beyond the scope of this report. Rather, to convey a sense of the type of results produced by this analysis, as well as potential pitfalls, we present in some detail the CCACF results for the local cattle price ratio using the 2-month SSWI index.

During the 2011 drought year (December 2010 to November 2011), a pattern of price ratio reductions that apparently lagged behind drought indices occurred in southern Somalia (Fig. 38). Locations where this reduction occurred may theoretically indicate where and when pastoralists begin to cull the herds that hydrologic conditions can no longer support, resulting in a local glut and consequent lower prices. Note that the same data can also indicate locations where, as conditions became wetter, the price ratio rose. A plot of the cross-correlation function values shows, for example, that the peak value of the CCACF in the Jamame market occurs at a lag of 0 months, indicating no lag (Fig. 39); this is the value that is plotted at the appropriate location on the map in Figure 38. The same is done for each market location where a statistically significant price ratio change occurred following a period of drought.

If we view the initial outputs (Fig. 38) of the CCACF analysis without further investigation, it might appear that the results are sound and that the behavior confirms anticipated price reductions of livestock following drought. However, a closer look at the data that produced this CCACF function (Fig. 38) shows that this result may be due in part to a mathematical artifact of the analysis. Because the largest values of both SSWI and the price ratio occur at the edge of the graph, then as the autocorrelation function is computed, the area under the increase in SSWI is not only decorrelated from the increase in the price ratio but is also no longer included in the computation of the CCACF. It is not immediately apparent whether the more important factor in this result is the actual effect of the arrival of the rains ending the drought or only the effect of the placement of those rains within the lag analysis of the time series. This analysis,

therefore, may not be appropriate for relatively short time series, such as that of a single year.

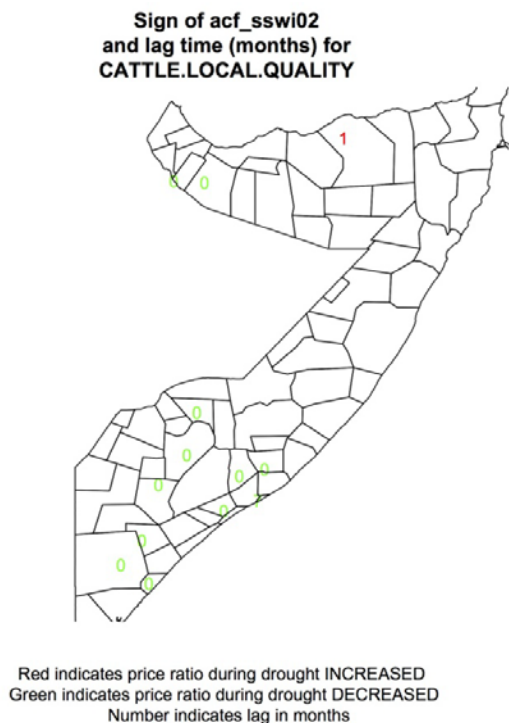


Figure 38. A CCACF analysis appears to demonstrate short-term effects of drought on cattle prices in southern Somalia during the 2011 drought year. However, this initial analysis may not be statistically robust due to edge effects of the CCACF analysis described below.

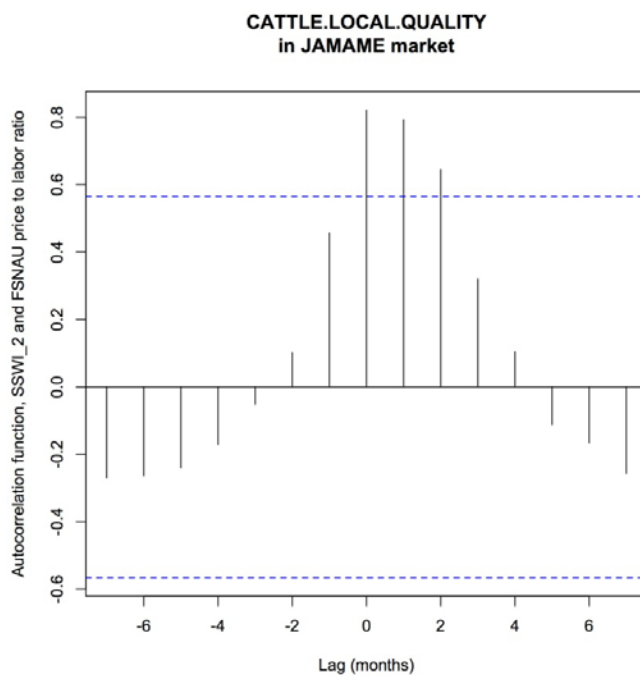


Figure 39. The cross-correlation function for local-quality cattle in one particular market peaks concurrently (lag = 0) with the 2-month SSWI index at that market location.

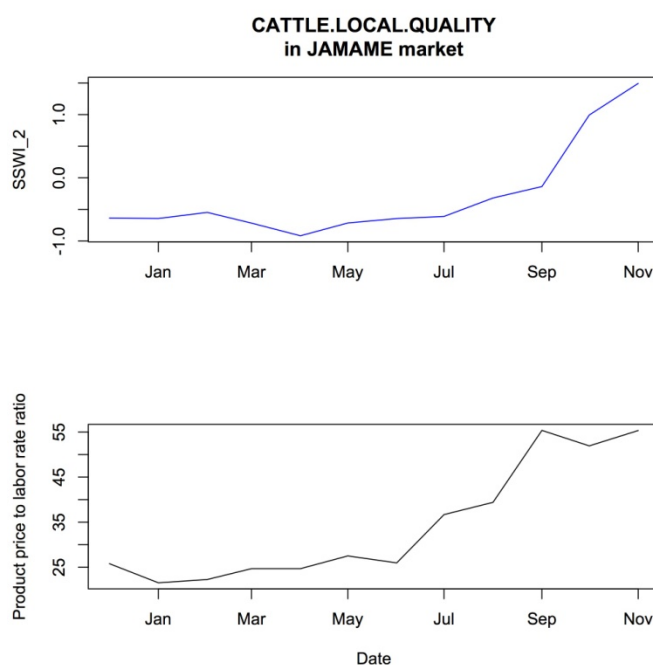


Figure 40. For the period covering the 2011 drought, the largest deviations from the mean occur near the ends of the time series. This has negative implications for the robustness of the CCACF analysis.

On the other hand, we also found that there are equally important potential problems with performing this analysis on a longer time series. In theory, a relatively short time series places too much weight on behavior at the temporal edges of the series, and an expanded time series would minimize the importance of edge effects and should produce more robust results. Figure 41 shows that the available data for this particular market product (local quality cattle in Jamame) is largely limited to the period from mid-2002 onwards, which would expand the time frame of the analysis considerably. However, whereas the shorter dataset from the drought period is likely to show economic fluctuations associated with the drought, enlarging the time period of study also enlarges the sphere of influence of non-drought-related externalities that affect livestock prices and day labor rates. For example, during the period between 2002 and 2012 when relatively consistent data was available, southern Somalia experienced an Ethiopian military intervention, the rise of the Islamic Courts Union, an al-Shabaab insurgency, and the lifting of a 9-year Saudi Arabian livestock import ban, each of which may have had significant effects on local cattle prices (whether raised for local consumption or for export) but none of which is directly related to drought conditions.

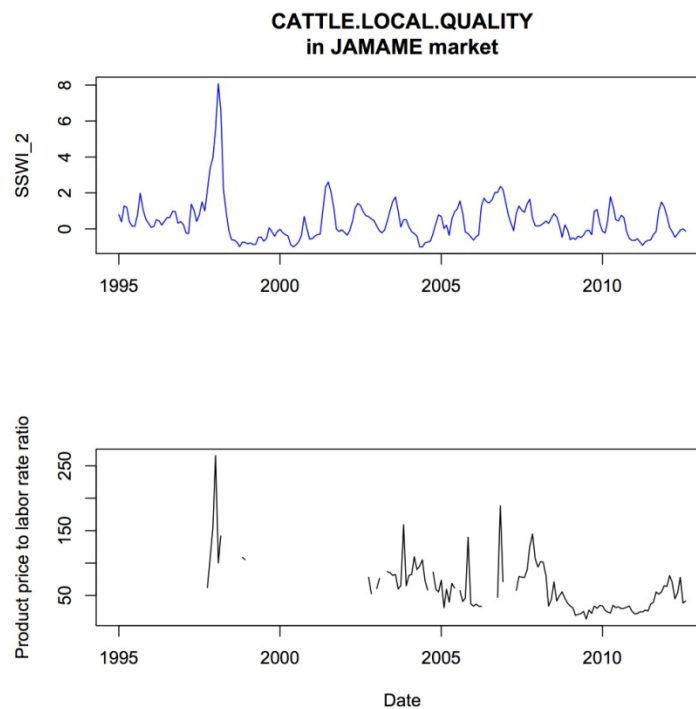


Figure 41. Raw data for the 2-month SSWI and FSNAU price-to-day-labor rate for local-quality cattle.

Maps of CCACF maxima produced using this expanded time period suggest that for both local- and export-quality cattle, there is a different price-ratio behavior for areas closer to the port city of Mogadishu than for the inland areas in the southwestern part of the country (Fig. 42 and 43). While these results have not been adjusted to remove the influence of other economic factors, they suggest that market access, in this case to a major port, may be a determining factor during periods of drought-induced economic stress.

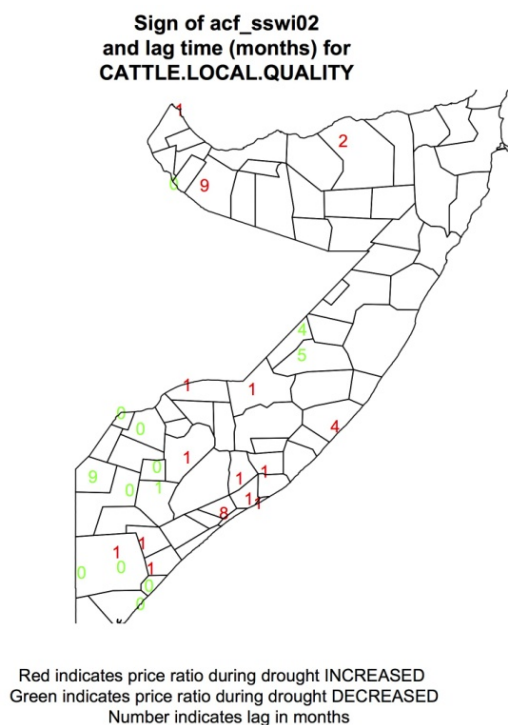


Figure 42. Price ratio response to the 2-month SSWI for local-quality cattle from July 2002 through August 2012. In markets closer to Mogadishu, price ratios increased during dry periods; but in markets further from the port city, price ratios decreased.

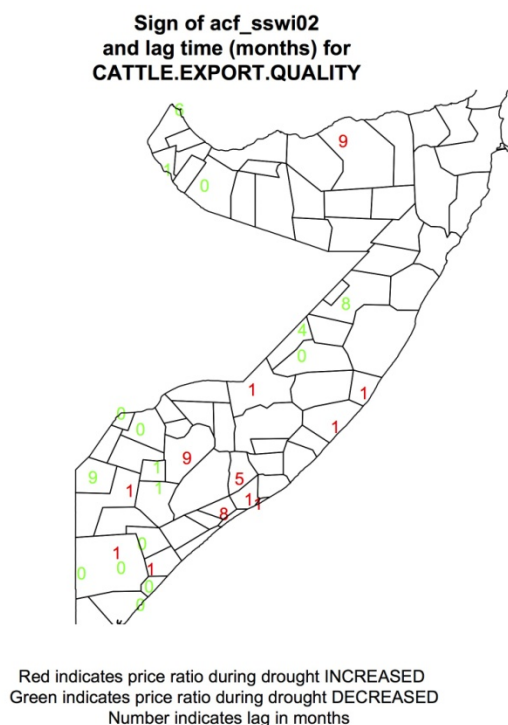


Figure 43. Price ratio response to 2-month SSWI for export-quality cattle from July 2002 through August 2012. The pattern of spatially-differentiated price ratios during drought is similar to that of local-quality cattle, but the variance in lag times is larger.

Overall, this type of analysis could be useful in a limited sense when queries aim to answer specific questions and the results consider a larger economic context. In future work, population movements that affect market demand could be considered; “market catchments” could be used rather than single grid cells to compute moisture indices against which lag times are calculated; the market-price-to-day-labor-rate ratio could be replaced with a market price in US dollars for products sold to external markets; effects of missing data values in the analysis could be considered; and multiple significant peaks in the CCACF could be retained rather than only the maximum. This present analysis, however, demonstrates some of the capabilities and limitations of this statistical technique in a spatial framework for analysis of climate-related economic effects.

3.4.4 Inter-market price correlations

Local markets typically do not exist in a vacuum only to redistribute goods amongst the local population but are themselves interconnected by physical proximity, preferred transportation routes, social and cultural connections, and multiple economic influences. The extent to which prices in one market behave similarly over time to prices in another market may therefore be an indicator of connectivity between these two markets, integrating and summarizing the various specific mechanisms for interconnection. Therefore, market prices could be used to highlight both particular locations that serve as hubs in a country-wide network and locations that are not well-connected and are potentially more vulnerable to local shocks. In addition, understanding how these patterns of connectivity change over time could potentially yield information on how market network structures change in response to external events, such as drought or conflict.

Graphs of the geographic correlation networks during the 2010–2011 year show that different products have distinct patterns of price correlations. On one side of the spectrum, for example, wheat prices exhibit only a small number of significant correlations between market locations. On the opposite side of the spectrum, the exchange rate between US dollars and Somali shillings exhibits a high degree of network connectivity (Fig. 44), demonstrating the fluidity of the monetary system.

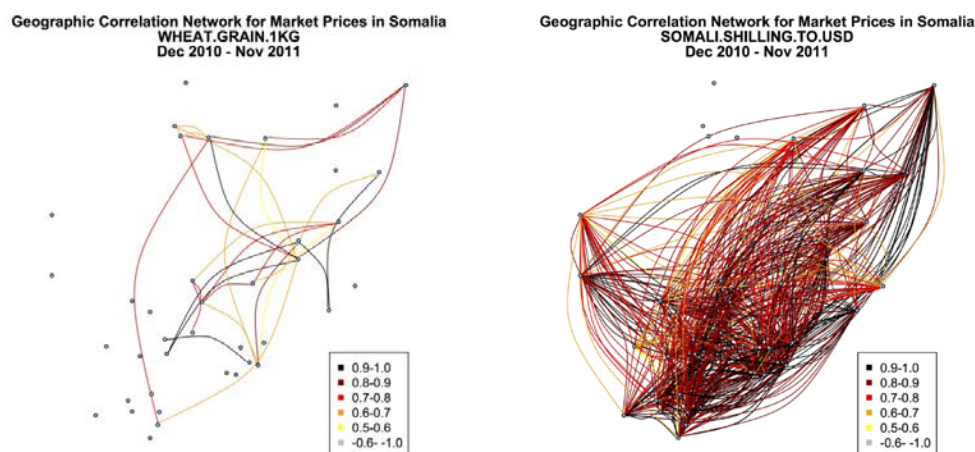


Figure 44. Market locations in Somalia are mapped as nodes on a network graph. Each graph edge connects two nodes that exhibit a statistically significant Pearson's r correlation of a minimum of 0.5, with darker lines indicating stronger correlations. The two nodes on the left-most side of these and subsequent graphs represent the two market locations whose geographic coordinates were unknown.

Network graphs of price correlations for staple cereals white maize and sorghum exhibit similarly changing patterns in the three years leading up to and encompassing the 2011 drought year. In 2008–2009, connections were widespread but not particularly strong, with similar stronger connections between some southern markets and Mogadishu-area markets (Fig. 45).

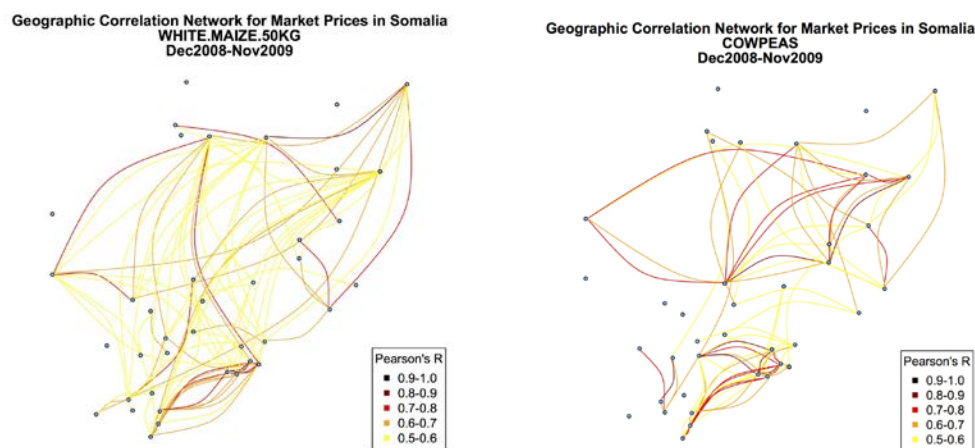


Figure 45. Correlation networks for white maize and cowpeas, December 2008–November 2009.

In 2009–2010, connections were more localized to southern Somalia with Mogadishu appearing as a particular hub for white maize (Fig. 44).

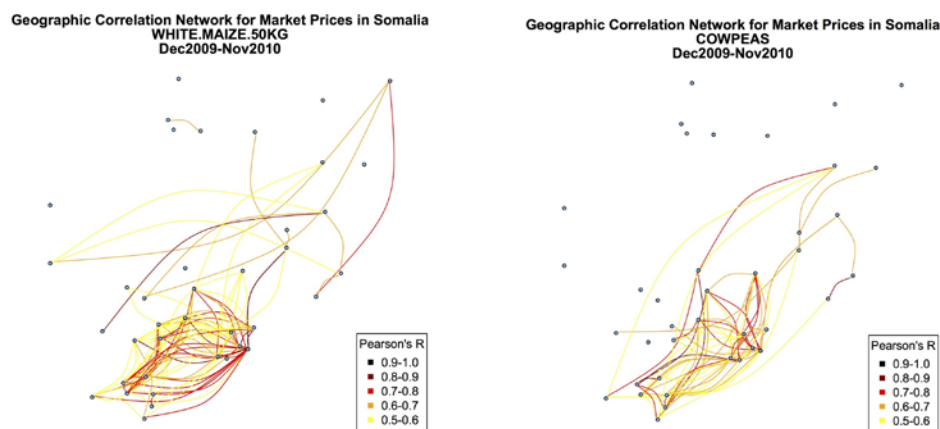


Figure 46. Correlation networks for white maize and cowpeas, December 2009–November 2010.

During the subsequent drought year, inter-market connections became much stronger, showing that market prices moved together over a wider area than during non-drought years (Fig. 47). In the case of maize, this is the likely result of poor yields in major surplus-producing regions, as depicted by an undated production and market flow map by FEWSnet for maize produced (Fig. 48). Whereas the extent of these stronger connections is largely limited to southern Somalia for maize, it reaches into northern Somalia for cowpeas. No similar map is provided by FEWSnet for cowpeas for comparison.

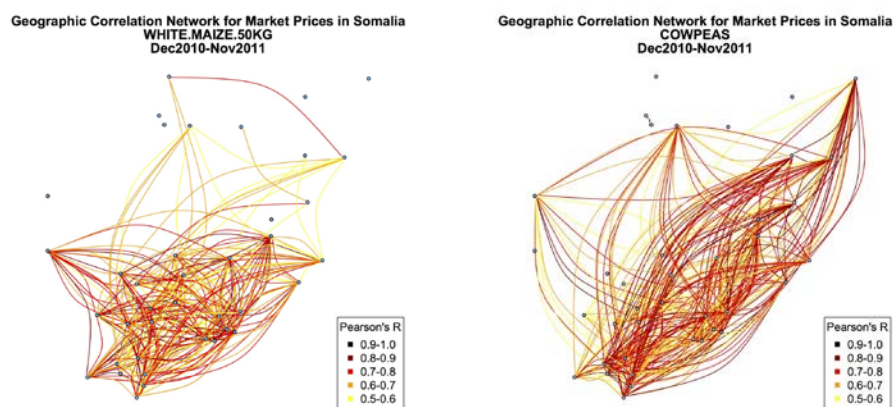


Figure 47. Correlation networks for white maize and cowpeas, December 2010–November 2011.

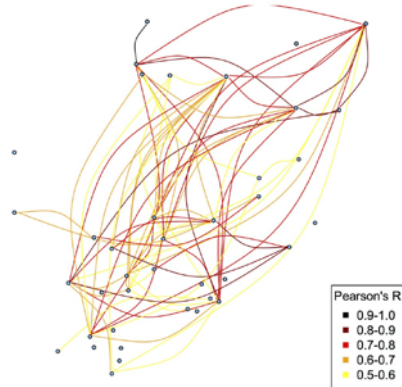


Figure 48. Production and market flow map for maize (FEWSnet 2008).

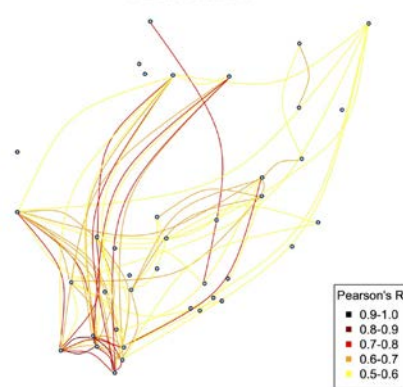
The same pattern of higher market connectivity during the drought year, however, is not found across all products. For example, network graphs of export-quality sheep and local-quality cattle change relatively little across these three years (Fig. 49). This time period also encompasses Saudi Arabia's lifting of a prolonged livestock import ban in 2009.

Qualitative analysis of the graphical outputs of a quantitative spatial correlation computation suggests that network analysis may in some cases contain and allow for the accurate display related to certain real effects on the ground. As these graphs become denser with better-connected markets, however, a simple visual analysis of the data may become inadequate and can be augmented by a quantitative summary, such as degree centrality.

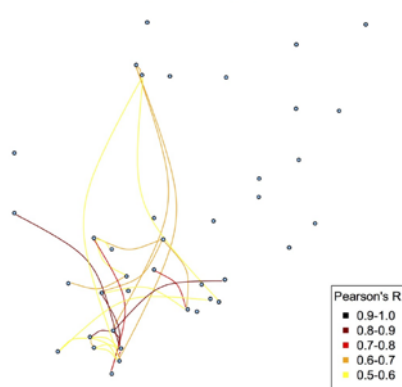
Geographic Correlation Network for Market Prices in Somalia
SHEEP.EXPORT.QUALITY
Dec2008-Nov2009



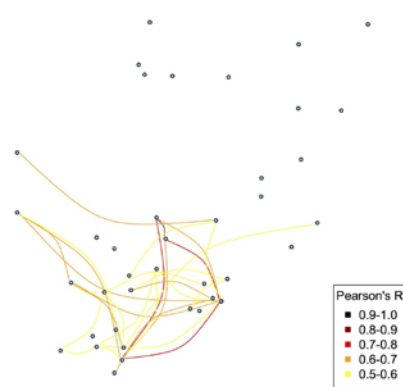
Geographic Correlation Network for Market Prices in Somalia
SHEEP.EXPORT.QUALITY
Dec2009-Nov2010



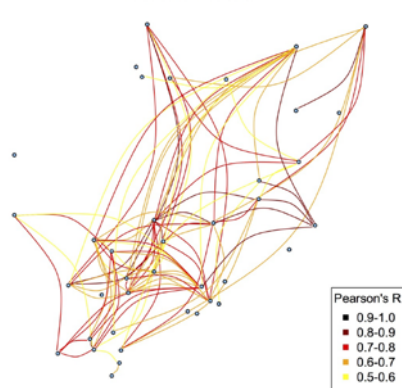
Geographic Correlation Network for Market Prices in Somalia
CATTLE.LOCAL.QUALITY
Dec2008-Nov2009



Geographic Correlation Network for Market Prices in Somalia
CATTLE.LOCAL.QUALITY
Dec2009-Nov2010



Geographic Correlation Network for Market Prices in Somalia
SHEEP.EXPORT.QUALITY
Dec2010-Nov2011



Geographic Correlation Network for Market Prices in Somalia
CATTLE.LOCAL.QUALITY
Dec2010-Nov2011

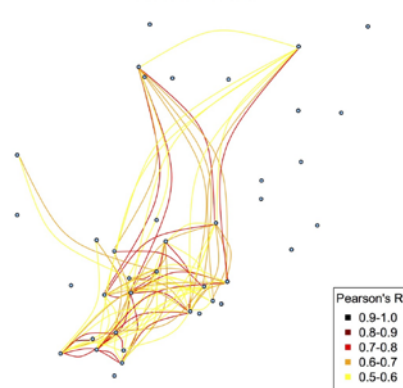


Figure 49. Correlation networks for export quality sheep and local quality camels, December 2008–November 2011.

3.4.5 Degree centrality

Degree centrality, the number of edges connected to each node in a network graph, is one measure of how well connected a particular network node is to other nodes in the network. In our analysis, we used degree centrality as a measure of how well correlated the price of a market product was to other prices in the country. The degree centrality was normalized across all products, showing which products tend to be relatively better connected than others. Figure 50 shows the degree centrality of a subset of products during the 2010–2011 drought year. As previously depicted in network graphs, the exchange rate between the US dollar and the Somali shilling exhibits very high connectivity. Livestock, the day labor rate, and sesame oil exhibit relatively low connectivity. White maize, reflecting the network depicted in Figure 47 for 2010–2011, shows relatively high connectivity in the south of the country and lower connectivity in the north. Figure 51 shows a summary connectivity metric across all products and markets, using the sum of all values of degree centrality for 2010–2011. Here, the results simply show that centrally-located markets in the south of Somalia are better connected than markets on the periphery of the country. We also note that Mogadishu has five separate markets assigned to the same point location in these network analyses and that superposition of these data points do not allow for all five values to be displayed.

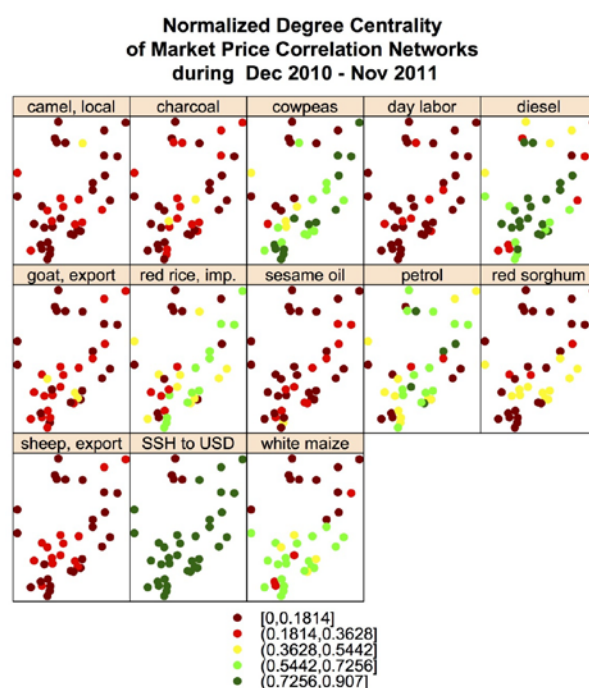


Figure 50. Normalized degree centrality for thirteen market products, December 2010–November 2011.

**Sum of Normalized Degree Centrality
of Market Price Correlation Networks
Dec 2010 - Nov 2011**

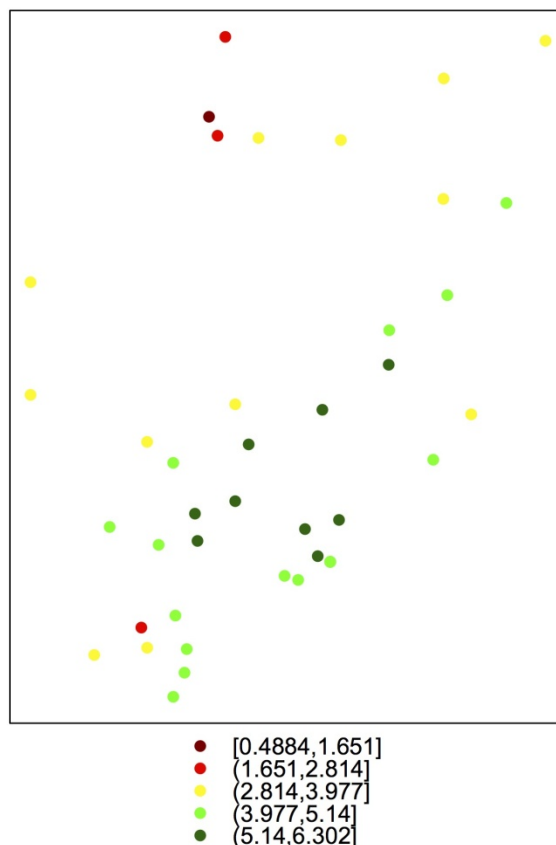


Figure 51. Normalized degree centrality summed across all market products, December 2010–November 2011.

If a market exhibits high degree centrality, it means that prices at that market moved in a similar pattern as prices in a larger number of other markets. This could indicate that the market is economically well-connected to those other markets and that trade is fluid between these locations. (The specific locations can be viewed using the full network graphs from Section 3.4.4.) It could also indicate that the same driver, such as drought, is affecting prices in multiple locations, irrespective of the degree of actual movement of goods between locations. If a market exhibits low degree centrality, it indicates that underlying conditions for pricing are dissimilar to that for most other markets. Flagging markets with unusual behavior as compared to others in the vicinity could potentially be useful for security analysis.

3.5 Census data

We used MERRA-Land-derived moisture indices to find areas that have historically been prone to drought and attempted to link publically available census data from Kenya to potential social and economic vulnerability to drought. We find that mapping the census data is useful, but is limited by the resolution at which the dataset is made available publically. The technique of association rules analysis, however, can overcome some of these limitations and could potentially be a simple and accessible tool to characterize certain aspects of a local community in a known geographic location.

3.5.1 Mapping of census data

Figure 52 presents the most frequent attributes of the selected variables from the Kenyan national census data for 1999. Approximately 70% of Kenyan households were classified as rural with the remaining 30% urban. Approximately 10% of the population lived in the district encompassing the capital city, Nairobi. Almost 70% of the population lived in owner-constructed houses, and similar numbers had no piped water supply with around 90% unconnected to a sewage system. Mapping the spatial extent of these and other variables gives a picture of the distribution of these infrastructure and household variables throughout the country.

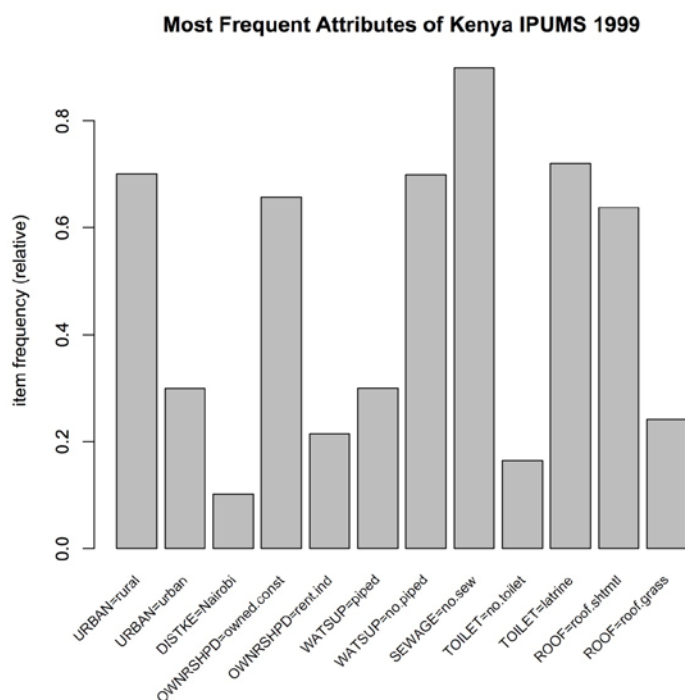


Figure 52. Most frequent attributes of Kenyan census data from 1999.

Maps presented below uniformly use the data and geospatial extents for 1989 as there were some changes in both district extents and in some of the selected census variables between the two census counts. Ideally, further use of geospatial census datasets should rely on official census boundaries rather than a mapping of census districts to administrative districts based on census district names alone.

Heavily urban districts were the south-central capital city, Nairobi, and southern coastal city of Mombasa (Fig. 53); and these urban households tended not to own their own houses (Fig. 54). Outside of these centers, both household ownership and owner-constructed houses are more common (Figs. 54 and 55). However, this renders finer distinctions among other types of household ownership and types of rental arrangements (Fig. 56) less easily distinguished using these maps. Considerable variability exists across the country in the fraction of households using piped water and different types of toilets; lack of a sewage system, however, is a fairly uniform characteristic (Fig. 57). Sheet metal and thatch are the primary roofing materials used, with sheet metal being common in the greater Nairobi area and thatch more common in northern and southern coastal regions (Fig. 58).

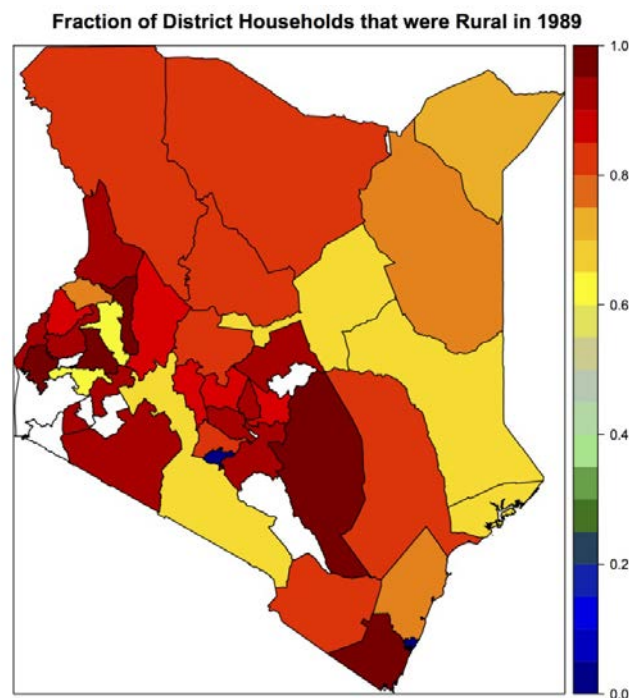


Figure 53. Rural household fraction.

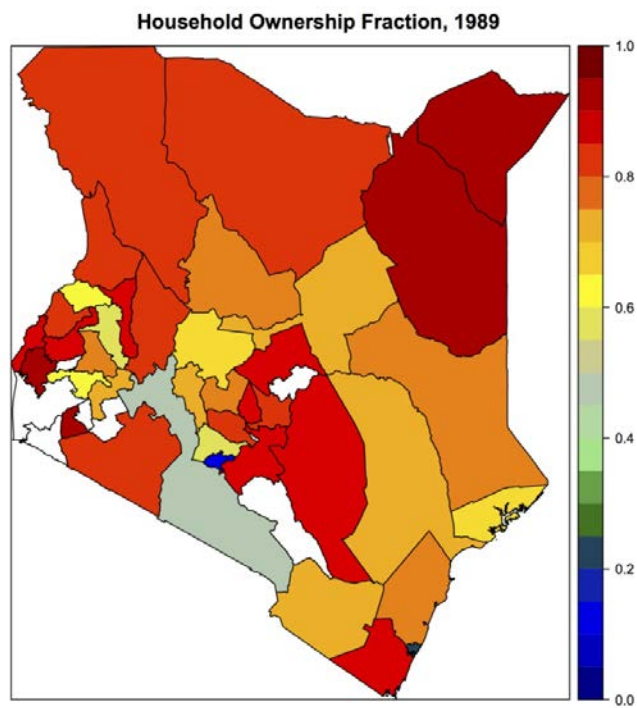


Figure 54. Household ownership fraction.

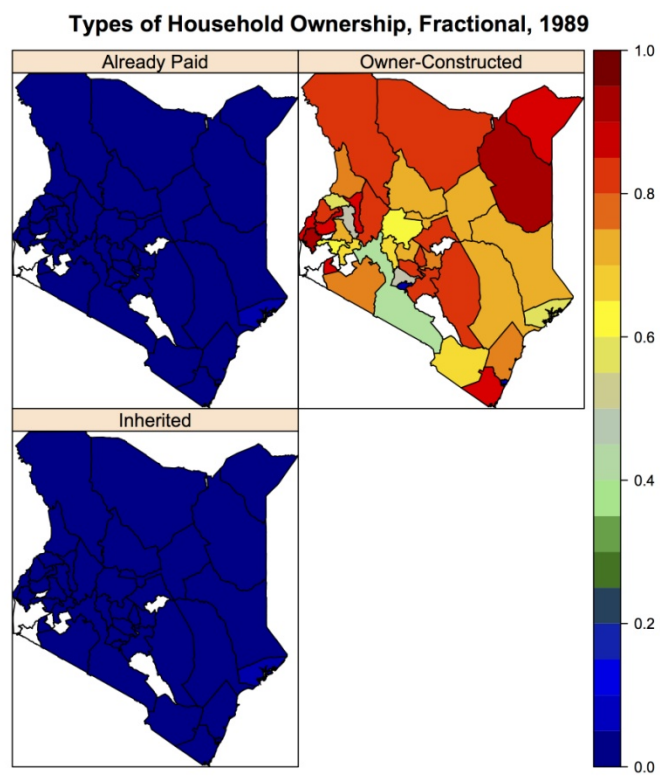


Figure 55. Household ownership categories.

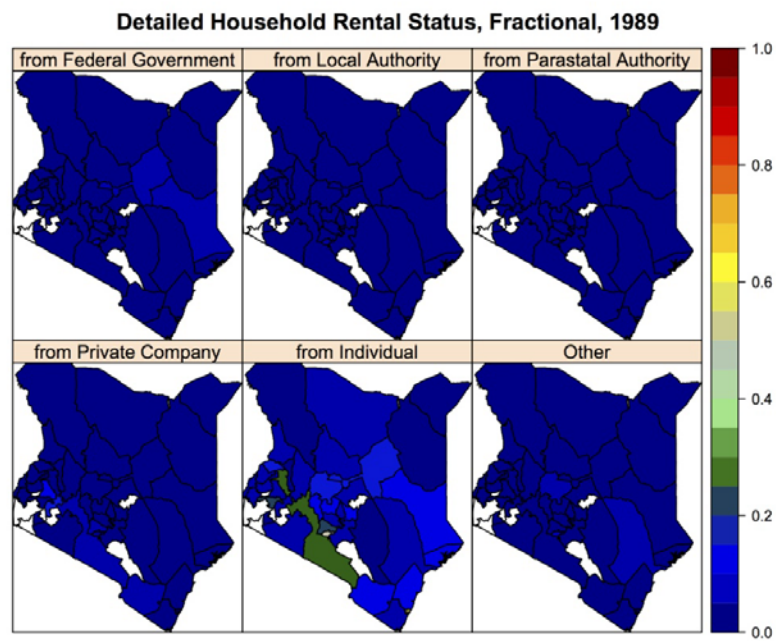


Figure 56. Household rental status.

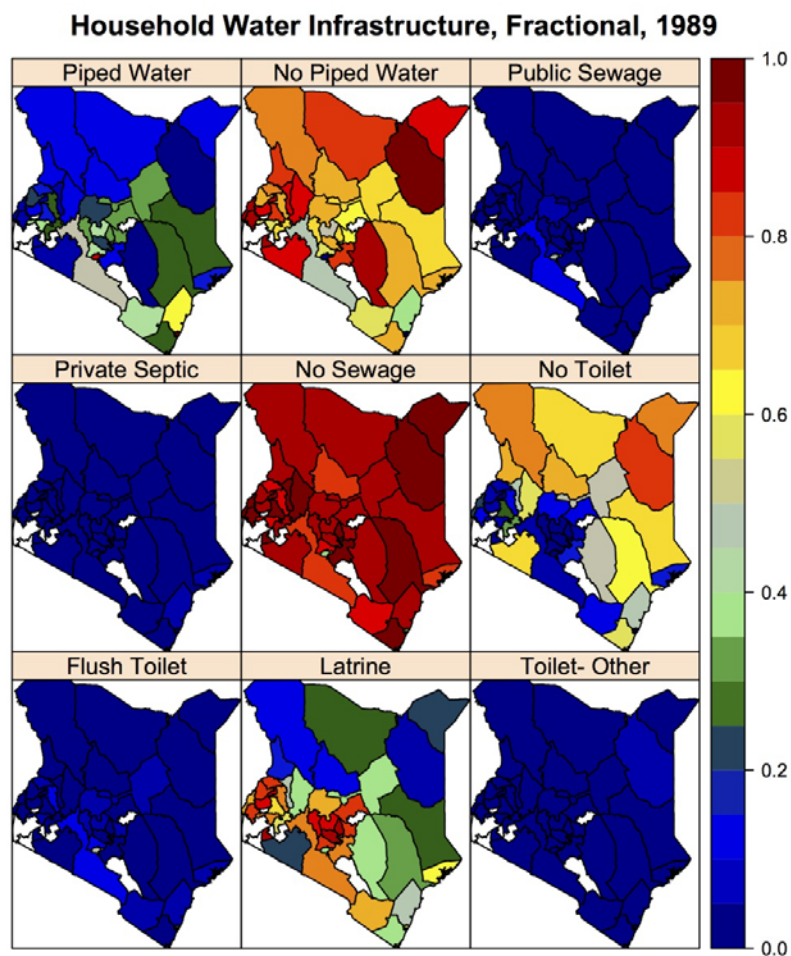


Figure 57. Household water infrastructure.

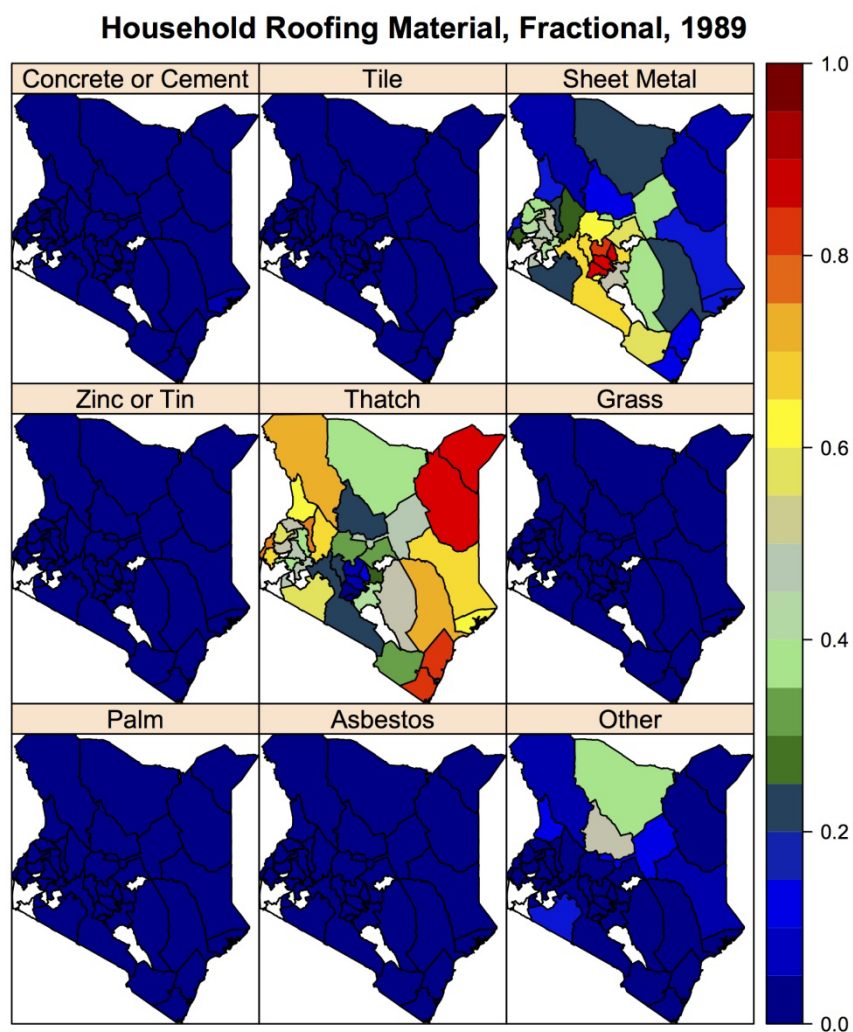


Figure 58. Household roofing material.

Although maps at this scale present an overview of general patterns seen in the census data that greatly improves upon country-wide statistics, they are of somewhat limited utility when attempting to narrow down relevant population characteristics of a specific community. For this purpose, association rules analysis is better suited.

3.5.2 Drought locations in Kenya

MERRA-Land-derived moisture indices calculated over Kenya indicate that the areas most subject to drought between December 2010 and November 2011 were the northern districts of Mandera, Marsabit, and Wajir, each of which had locations that experienced, for example, a 3-month SPI of -2 or lower during at least four consecutive months in this time period (Fig. 59). Other districts affected by drought to a lesser degree (3-month

SPI of -1 to -2) included Isolo, Garissa, Tana River, Lamu, Kilifi, Kwale, Mombasa, and Taita Taveta.

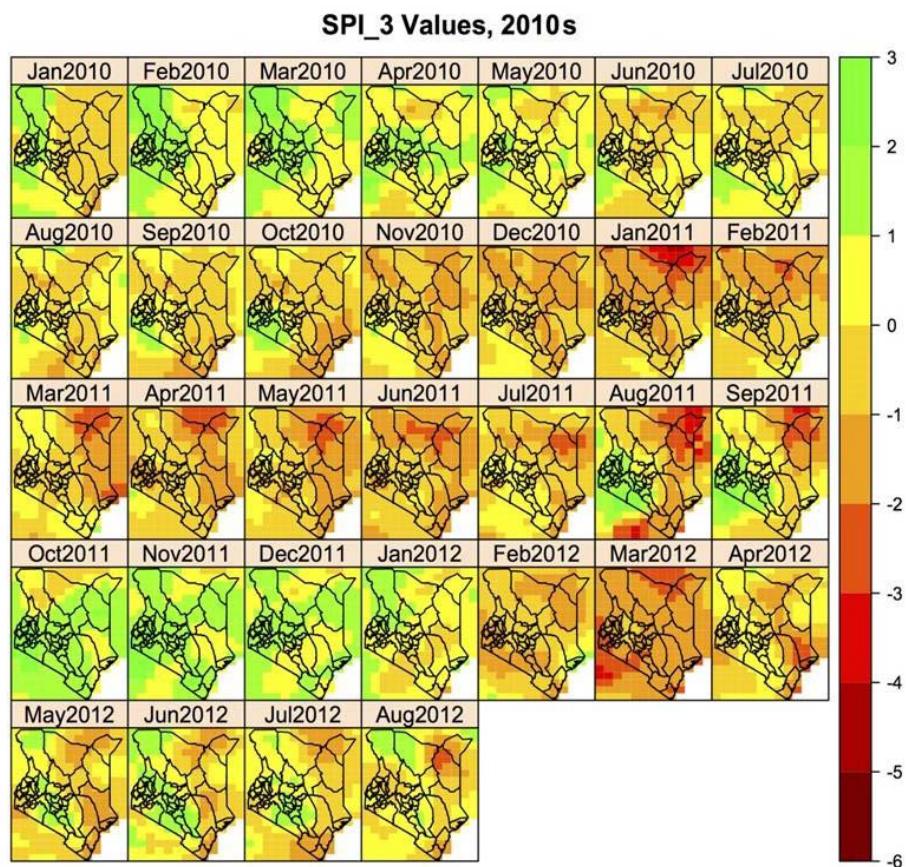


Figure 59. Three-month SPI across Kenya, January 2010–August 2012.

3.5.3 Association rules analysis

This example presents a simple methodology to infer the probability of certain unobservable census variables based on other observable variables and to explore further associations. Census data could potentially contain indicators of the likelihood of population movement in response to unmet water requirements. In particular, household water infrastructure may provide one source of information about a population's immediate vulnerability to the effects of drought. The 1999 census data represents the most recent census whose complete set of data was publically available during the course of this study. We used association rules analysis to perform queries on data related to household water infrastructure in the districts most affected by the 2011 drought. From the initial graphical depiction of 1989 census variables, we expect that the majority of dwellings in these areas were largely rural and owner-constructed and that households typi-

cally had neither piped water, sewage, nor toilet facilities and had thatch roofs (Figs. 55 to 58). Using association rules analysis, however, can help to bring into focus more specific combinations of these census variables. These in turn, with the assistance of tools such as high-resolution remote sensing to identify roof types, could be used assess the likelihood that specific communities had access to certain types of water infrastructure.

The set of queries posed via association rules analysis here focuses on just one district affected by drought, but the method can be easily applied to any other administrative unit within the dataset. The first query lists variables associated with rural households in the Marsabit district. We derived an initial set of rules from the country-wide dataset by setting the minimum support very low (1×10^{-6}) to include the low populations of this district. The minimum confidence was set to 1%, and the maximum rule length was set to three to include the two antecedent variables “Marsabit district” and “rural” and to produce one consequent variable. The Apriori rule mining algorithm produced a list of over 38,000 rules countrywide. These were then subset to list rules applying to only the rural households in the district in question. The minimum support for the initial rule generation was varied from 1×10^{-4} to 1×10^{-8} , and the results were found to change only minimally with one rule being removed from the final list of twelve when the support requirement was increased to 1×10^{-4} .

As listed in Table 4, the results show that for sampled rural households in Marsabit, 99.6% had no sewage, 92.9% had no piped water, and 85.9% had no toilet. 12.9% of households had a latrine, and 7% of households had piped water. 92.3% were owner constructed, and 2.4% were rented from an individual. 51% of households had a roof type listed as “other” while 20.8% had palm and 15.6% had grass roofs. A high lift of 25 for the rule associating a roof type of “other” with rural households in Marsabit indicates a particularly strong association with this category, which was not widely used outside of the district and may represent district-level differences in census questionnaires. Contrary to expected results based on 1989 census maps, thatch roofs did not appear on the list of rules with a minimum confidence of 1% for rural households in Marsabit; other census variables not on the list can also be presumed to apply to less than 1% of the rural Marsabit population. Results of the association rule analysis for this first query are summarized below.

Table 4. Association rules for rural households in the Marsabit district.

Antecedent variables	→ Consequent variables	Support	Confidence	Lift
{URBAN=rural, DISTKE=Marsabit}	{SEWAGE=no.sew}	4.25E-03	0.99556869	1.107419
	{WATSUP=no.piped}	3.97E-03	0.92909897	1.3296575
	{OWNRSHPD=owned.const}	3.95E-03	0.9239291	1.4062284
	{TOILET=no.toilet}	3.67E-03	0.85893648	5.2160924
	{ROOF=roof.other}	2.18E-03	0.50960118	24.2493386
	{ROOF=roof.palm}	8.86E-04	0.20753323	5.5256116
	{ROOF=roof.grass}	6.65E-04	0.15583456	0.6454555
	{TOILET=latrine}	5.49E-04	0.12850812	0.1783822
	{ROOF=roof.shtmtl}	5.08E-04	0.11890694	0.1864746
	{WATSUP=piped}	3.00E-04	0.07016248	0.2337688
	{OWNRSHPD=rent.ind}	1.01E-04	0.02363368	0.1103592
	{OWNRSHPD=owned.pd}	9.78E-05	0.02289513	0.6017556

To identify associations linked to urban households in Marsabit, the second query repeated a similar process (Table 5). Like rural residents, 98.6% of urban households had no sewage system; but unlike the rural households, only 33.8% had no piped water, and only 2.9% had no toilet. 95% had a latrine, 66% had piped water, and 1.4% had a flush toilet. The vast majority of households had sheet metal roofs (97.8%), and 2.1% had asbestos roofs. Only 41.0% of household dwellings were owner-constructed, 36.0% were rented from an individual, and 10.0% were rented from the government. 5.7% had been purchased, and 3.6% were owned via inheritance with smaller percentages representing rental from private companies and parastatal organizations.

Because the highest lift of this group was for the rule associating renting from the government with households in urban Marsabit, we conducted a further query investigating associations with this particular combination of census variables. A new rule set was derived with a maximum rule length of four to allow for the sorting of rules by three known variables (Marsabit district, urban, and government rental) to reveal other associations. The rules (Table 6) show that all of these properties have latrines and that none have sewage systems, 92.8% have piped water to the households and sheet metal roofs, and 7.1% have asbestos roofs and do not have piped water. A further query on urban Marsabit households that rent from the government and have asbestos roofs shows, however, that these last two variables do not overlap and that all of the asbestos-roof households

have piped water (Table 7). If the same analysis is performed on Garissa households, for example, we find that all urban households that rent from the government and have asbestos roofs have piped water as well as a flush toilet and a public sewer.

Table 5. Association rules for urban households in the Marsabit district.

Antecedent variables	➔ Consequent variables	Support	Confidence	Lift
{URBAN=urban, DISTKE=Marsabit}	{SEWAGE=no.sew}	4.32E-04	0.985612	1.096343
	{ROOF=roof.shtmtl}	4.29E-04	0.978417	1.534393
	{TOILET=latrine}	4.16E-04	0.94964	1.318196
	{WATSUP=piped}	2.90E-04	0.661871	2.205234
	{OWNRSHPD=owned.const}	1.80E-04	0.410072	0.624133
	{OWNRSHPD=rent.ind}	1.58E-04	0.359712	1.679702
	{WATSUP=no.piped}	1.48E-04	0.33813	0.483906
	{OWNRSHPD=rent.gov}	4.41E-05	0.100719	5.652873
	{OWNRSHPD=owned.pd}	2.52E-05	0.057554	1.512698
	{OWNRSHPD=owned.inh}	1.58E-05	0.035971	1.767383
	{TOILET=no.toilet}	1.26E-05	0.028777	0.174755
	{ROOF=roof.asb}	9.46E-06	0.021583	1.720034
	{OWNRSHPD=rent.pc}	9.46E-06	0.021583	0.679847
	{OWNRSHPD=rent.ps}	6.31E-06	0.014388	1.45819
	{TOILET=flush.toilet}	6.31E-06	0.014388	0.144047

Table 6. Association rules for urban Marsabit households that rent from the government

Antecedent variables	➔ Consequent variables	Support	Confidence	Lift
{URBAN=urban, DISTKE=Marsabit, OWNRSHPD=rent.gov}	{TOILET=latrine}	4.41E-05	1	1.388101
	{SEWAGE=no.sew}	4.41E-05	1	1.112348
	{WATSUP=piped}	4.10E-05	0.928571	3.093833
	{ROOF=roof.shtmtl}	4.10E-05	0.928571	1.456223
	{ROOF=roof.asb}	3.15E-06	0.071429	5.692493
	{WATSUP=no.piped}	3.15E-06	0.071429	0.102223

Table 7. Association rules for urban Marsabit households that rent from the government and have asbestos roofs

Antecedent variables	➔ Consequent variables	Support	Confidence	Lift
{URBAN=urban, DISTKE=Marsabit, OWNRSHPD=rent.gov, ROOF=roof.asb}	{WATSUP=piped}	3.15E-06	1	3.33182
	{TOILET=latrine}	3.15E-06	1	1.3881
	{SEWAGE=no.sew}	3.15E-06	1	1.112348

To the extent that household water infrastructure is a characteristic of sedentarization as opposed to transhumant or nomadic populations, observing changes in water infrastructure over time through census data could indicate the extent to which transhumance continues to provide adequate livelihoods. This analysis, however, does not presume to assess the extent to which household water infrastructure is actually indicative of drought resilience. If a certain condition, in this case having piped water, could be linked to a certain resilience to the initial effects of a drought, then the following conclusion could result from this analysis: 100% of households with tile roofs in Marsabit would be subject to displacement, as would 96% of households with palm roofs, 88% with grass roofs, 75% with asbestos roofs, and 59% with sheet metal roofs.

While the need for permanent household water infrastructure may be associated with sedentarization, use of piped water supply as an indicator misses critical information related to potential population response to drought. In particular, it does not provide indicators of the quantity, quality, or seasonal stability of the water supply that arrives via pipe, nor does it differentiate between actual sources of piped water, which may also be available to households via transport mechanisms other than a direct household line. In addition, the advantages in the amount of detail typically available in census data is somewhat offset by the low temporal resolution of that data, which is typically on a scale of a decade. In the present case of Kenya, raw data from a census more recent than 1999 had not yet been made publically available at the time of the study. In summary, the method discussed here provides an accessible mechanism with which researchers can access and disaggregate census datasets and could potentially be useful in remotely assessing characteristics of specific communities based on a smaller number of observed variables; however, the data that was collected on household water infrastructure as part of the Kenyan 1989 and 1999 census may not be adequate alone to characterize potential drought response on the part of the households surveyed.

4 Conclusions and Recommendations

This study aimed to quantitatively compare several existing independently-developed geospatial datasets to improve understanding of conditions under which drought might be associated with socio-economic outcomes of potential interest to US security. We first discuss the strength and mechanism of the potential climate-security link and then below summarize conclusions from the six components of the study.

Data on internally displaced persons in southern Somalia suggest that there may have been a significant if indirect connection between climate extremes and insecurity following the 2011 drought. The population movement data support observations made by others that the reigning governing body (al-Shabaab) suffered a loss of popular support due to its inability to provide for people during the drought. Aid access was limited by different but related mechanisms: by the local regime itself, indirectly by the security situation, and also by international anti-terrorism or anti-corruption legislation preventing aid delivery via certain pathways. In turn, the resulting lack of an effective response to the drought may have been one of several factors that led to al-Shabaab's ousting from large parts of southern Somalia in the following year (Figs. 10 and 11). This data is preliminary to the current study rather than a result of it but nonetheless demonstrates the role of empirical observations in addressing these types of questions. Note that although the numbers of insecurity-induced movements are an order of magnitude higher than the drought-induced movements, if there is a causal connection then the drought-induced movements may have an outsized importance.

Any such climate-security link invites open discussion surrounding the possibility that climate situations could therefore be used “strategically” by multiple actors and that there is likely to be an inherent conflict of interest between humanitarian actors wishing to respond to affected persons in the short term and military and political actors wishing to shape public opinion by limiting the capability of certain governing groups to respond to climate-related stress.

4.1 MERRA-Land

The MERRA-Land data adequately covered the needs of the study: we could obtain consistently-produced climate reanalysis globally at a scale reasonably compatible with most of the socio-economic datasets used in the study, and the MERRA-Land output presented a long enough record for us to attempt historical comparisons of seasonal moisture indices. While the climate datasets were not themselves evaluated as part of this project, results in certain locations (e.g., root-zone soil water in northeastern Somalia in November 2004) might benefit from a closer inspection to determine if some error or model instability may have been the source of apparent anomalies. Use of retrospective analyses underscores how investment in robust geophysical observation networks has payoffs that last well into the future.

In future work where suitable in situ observations are unavailable, MERRA-Land outputs could be compared with other potential drought indicators available in the study area. For example, NDVI anomalies from MODIS-derived imagery or rainfall anomalies from TRMM-derived datasets could potentially provide a way to assess the overall validity of MERRA-Land precipitation data in the study region despite the shorter timescales of these remotely-sensed records. The methods our study used to ingest the climate data can be reused in future processing of results from new model runs, for example, that include assimilation of remotely-sensed soil moisture or groundwater.

4.2 SPI/SSWI indices

Multiscale standardized indices used to investigate the time scales associated with socio-economic phenomena reinforce the importance of cumulative climate impacts (i.e., the effects of not just one but two weak or failed bi-annual rainy seasons and the effects of a series of several years with anomalously low rainfall despite the presence of one year of unusually strong rains). The methods used in this study could be easily extended for use with other climate datasets and phenomena (e.g., for floods rather than droughts) and to other spatial and temporal scales of interest.

The indices used here are not without some inherent limitations. As time scales of an investigation increase, so does the potential influence of other unquantified, non-climatic factors, such as changes in governance or external economic drivers, that may also influence results and prevent inde-

pendent analysis of climate factors alone. Further, it may be advisable to include a parallel type of drought index that is based on absolute rather than relative historical values. The SPI/SSWI-type index is useful as it will not confuse anomalous dry seasons with true droughts, but it may not adequately distinguish absolute differences, such as that of arid areas from non-arid areas.

4.3 Drought and PMT

Mapping of correlation values between moisture indices and drought-induced population movements in Somalia from 2008 to 2012 shows that the locations at which correlations are strongest are not randomly distributed. Rather, they are concentrated in the southern and southwestern parts of Somalia, particularly for cumulative indices at time scales of 6–18 months and again for the longer 36- and 48-month time scales. Irrespective of the strength of the actual correlation coefficients, results are significant for large areas of non-coastal southern Somalia at a variety of time scales. Areas where correlations are not significant may also yield useful information: for example, on the shorter time scales of up to 6 months, an area around Mogadishu showed consistently insignificant results in the correlation analysis (Fig. 22). This could potentially be an indicator of the de facto extent of the reach of aid and commerce to populations not dependent on subsistence agriculture or of the population expecting aid and commerce around the port city would provide for them during the drought.

In the northern half of the country, drought indices were only sporadically correlated with population movements. The differences between responses in northern and southern Somalia suggest there are significant geospatial factors of interest that are influencing population response. Northern Somalia differs from southern Somalia in at least two significant respects: its climate is routinely arid and so moisture anomalies are less disruptive to the economy, and it was not governed by al-Shabaab and therefore was not subject to restrictions on aid associated with insecurity or anti-terrorism legislation. A separate study on seasonality patterns in Africa (Herrmann and Mohr 2011) sheds light on these possibilities. The differences in seasonality classes found between northern and southern Somalia could indicate that (1) seasonal moisture anomalies used in the current study are ineffective for understanding population movement response to drought in arid regions and that a drought indicator that took seasonality classes into account may provide additional useful distinctions in this type of analysis

or (2) there may be underlying social and economic factors that limited the extent of al-Shabaab influence to the regions of southern Somalia in which two unimodal wet seasons are the norm, suggesting the possibility of a more complex historical or anthropological relationship between climate and the evolution of governance structures. Both of these possibilities point to areas of research that may be of interest to the intelligence and defense communities, with the first possibility being amenable to a data-driven approach similar to the one used in this study.

At the shortest time scales used in the study, that of one month, almost no correlations were significant, particularly for the standardized precipitation index. This may indicate either that population response to drought is not immediate or that finer temporal resolution (e.g., daily) is needed to look at shorter-term movements, both for IDP movement as well as climate data. In general, aside from the first month, differences between the indices based on precipitation and those based on soil water moisture did not produce particularly different results, and future studies may not require the soil water moisture index at all.

The analysis that considered livelihood-zone and seasonal calendar distinctions between geographic areas suggests that use of averaged shorter-term precipitation-based indices might provide a useful way to estimate potential population movement response to drought in areas where a governance situation prevents aid access as occurred in southern Somalia. The fact that one version of the non-weighted index produced high confidence levels suggests that the livelihood-zone distinction, however, is not essential. To the extent that longer-term moisture indices may be more predictable than shorter-term ones, use of livelihood-zone-weighted indices may still provide some benefits. We note that the use of the seasonal calendars in this analysis inherently favors SPI rather than SSWI as the seasonal calendars demarcate rainy seasons and not seasons of favorable soil moisture, which would tend to continue beyond the end of a rainy season. Linear models can be run in “predictive” mode and confidence intervals can be calculated that may be narrow enough to make these estimations useful under certain conditions; however, any model applicability to new regions would require careful application as the complex situation in Somalia should be presumed to be fairly unique.

It is likely that mobile-device-derived location datasets could significantly expand the reach of PMT analyses to other locations where on-the-ground

surveying has not been conducted (Lu et al. 2012). However, this would present its own set of challenges as reasons for movement could not be easily separated except in cases with obvious causes. Application of this research may prove to be most useful, in fact, in attempts to separate out climate-induced population movements from an otherwise undifferentiated population movement dataset.

4.4 Drought and FSNAU market prices

Correlation analysis between moisture indices and price-to-day-labor ratios found that the strongest relationships occurred for white maize and red sorghum in southern Somalia during the December 2010–November 2011 period. The strongest negative correlations peaked near 8-month indices and again near longer-duration indices, such as the 48-month SPI. These price ratio increases during drought for these two staple products appear to be broadly consistent with patterns seen in previous drought-population movement correlation data.

Maps of CCACF maxima suggest that for some livestock products, there is a different price ratio behavior for areas closer to the port city of Mogadishu than for the inland areas in the southwestern part of the country. While these results have not been adjusted to remove the influence of other economic factors, they provide evidence that market access may be a determining factor in price movements during periods of drought-induced economic stress.

A pattern of increasing market price correlations during the drought year was found for some but not all market products. During the drought year, intermarket connections became much stronger for some staple cereals, showing that market prices moved together over a wider area than during non-drought years. In the case of maize, this is likely simply the result of poor yields in major surplus-producing regions. Network graphs of export-quality sheep and local-quality cattle change relatively little across the same time period, however.

Qualitative analysis of the graphical outputs of a quantitative spatial correlation computation suggests that network analysis techniques may unearth and display certain real effects. It is not clear from this research whether the results produced offer opportunities to improve upon existing knowledge or the extent to which security analysts could adequately separate out the mechanisms that cause the observed effects.

4.5 Drought and Kenyan census data

The method of association rules analysis discussed in this paper provides a mechanism to effectively disaggregate census datasets that could potentially be useful in remotely assessing characteristics of specific communities based on a smaller number of known or observed variables. The data used in this study related to household water infrastructure collected as part of the Kenyan 1989 and 1999 censuses is not adequate in and of itself to characterize potential drought response on the part of the households surveyed. However, the technique of associate rule analysis is both accessible and powerful. Including other census variables, such as household income sources, in this type of analysis may improve understanding of potential population response; and the technique could easily be applied in a variety of situations for on-the-fly community social, health, and infrastructure assessments in areas where historical census data exist. Census data in many countries, however, is justifiably subject to access restrictions for certain variables and does not produce temporally high-resolution data.

4.6 *R* and open-source libraries

In this study, we successfully used *R* to develop a set of flexible and extensible techniques for climate and socio-economic data ingestion and processing for potential use in future analyses. Because *R* is open-source, researchers in many disparate fields have used it to develop and publish methods for processing datasets commonly found in their fields. By providing code in one language that makes these widely varying types of data accessible and compatible, research that attempts to combine information from such widely varying fields of study as atmospheric physics and social science can find a practical common operating platform. Furthermore, as all code is open-source, other researchers have free access to the functions used in this analysis to perform reanalysis of the methods used in the current study. This enhances transparency and allows for modifications to the original code to suit study needs as they change.

The development of code to bring together disparate datasets in some type of robust analysis should not be presumed to be trivial or treated as an afterthought nor should it be presumed that individually robust data sources such as those used in this study can simply be “shaken together” in some analytic way to produce useful results. However, attempts to combine pre-existing data sources in new and potentially useful ways should not be ne-

glected. Robust data sources often require a significant amount of investment, and it is logical to attempt to maximize returns on these existing investments. Exploration of existing data sources can also help guide the appropriate collection of new datasets that are more targeted to address specific concerns of the intelligence and defense communities.

References

- Agrawal, R., T. Imielinski, and A. Swami. 1993. Mining association rules between sets of items in large datasets. *SIGMOD Record* 22(2):207–216.
- Agrawal, R., and R. Srikant. 1994. Fast Algorithms for Mining Association Rules. In *Proceedings of the 20th International Conference on Very Large Databases, VLDB*, 487–499.
- Alcorn, J., J. Bitting, K. Canada, K. Herrmann, R. Lueken, J. Ruffing, and K. Siroonian. 2011. *Environmental Change and Fragile States—Early Warning Needs, Opportunities, and Intervention*. Fort Belvoir, VA: Army Environmental Policy Institute.
- Almquist, Z. W. 2010. US Census Spatial and Demographic Data in R: The UScensus2000 Suite of Packages. *Journal of Statistical Software* 37(6).
- Army Research Office. 2011. *Minerva Research Initiative*. Research Triangle Park, NC: Army Research Office.
- Borgelt, C. 2003. Efficient implementations of apriori and eclat. In *Proceedings of the 1st IEEE ICDM Workshop on Frequent Itemset Mining Implementations, Melbourne, FL*.
- Bradley, J. V. 1968. *Distribution-Free Statistical Tests*. Englewood Cliffs, NJ: Prentice-Hall.
- Brin, S., R. Motwani, J. D. Ullman, and S. Tsur. 1997. Dynamic itemset counting and implication rules for market basket data. *SIGMOD Record* 26(2):255–264.
- Chatterjee, S., and A. S. Hadi. 1986. Influential observations, high leverage points, and outliers in linear regression. *Statistical Science* 1(3):379–393.
- Checchi, F., and W. C. Robinson. 2013. *Mortality among populations of southern and central Somalia affected by severe food insecurity and famine during 2010–2012*. Washington, DC: Food and Agriculture Organization and Famine Early Warning Systems Network.
- Department of State and US Agency for International Development (USAID). 2010. *Leading Through Civilian Power: The First Quadrennial Diplomacy and Development Review*. Washington, DC: Department of State.
- Dinku, T., P. Ceccato, E. Grover-Kopec, M. Lemma, S. J. Connor, and C. F. Ropelewski. 2007. Validation of satellite rainfall products of East Africa's complex topography. *International Journal of Remote Sensing* 28(7).
- Edwards, D. C., and T. B. McKee. 1997. *Climatology Report 97-2: Characteristics of 20th century drought in the United States at multiple time scales*. Fort Collins, CO: Department of Atmospheric Science, Colorado State University.

- FEWSnet. 2008. Production and Market Flow Maps: Somalia Maize First Season. *Famine Early Warning System Network*. http://www.fews.net/sites/default/files/documents/reports/so_fullmap_maize_s1.pdf.
- FEWSnet. 2011. Somalia Livelihoods Zone Map. *Famine Early Warning System Network*. <http://www.fews.net/east-africa/somalia/livelihood-zone-map/wed-2011-07-20>.
- FEWSnet. 2013. Somali Seasonal Calendar. *Famine Early Warning System Network*. <http://www.fews.net/east-africa/somalia/seasonal-calendar/tue-2013-12-17>.
- FSNAU. 2013. *Integrated Database System*. <http://www.fsnau.org/ids/>.
- FSNAU-Somalia. 2013 *FSNAU Technical Series*. Nairobi, Kenya: Food Security and Nutrition Analysis Unit–Somalia. <http://www.fsnau.org/products/technical-series> (accessed 30 January 2014).
- Gemenne, F. 2011. Why the numbers don't add up: A review of estimates and predictions of people displaced by environmental changes. *Global Environmental Change* 21:S41–S49.
- GeoNames. n.d. *GeoNames*. <http://www.geonames.org> (accessed 27 February 2013).
- Goddard Earth Sciences Data and Information services Center (GES DISC). 2010. MDISC Data Subset. *MDISC Modeling and Assimilation*. Greenbelt, MD: NASA. <http://disc.sci.gsfc.nasa.gov/daac-bin/FTPSubset.pl>.
- Hahsler, M., B. Bruen, and K. Hornik. 2012. arules: Mining Association Rules and Frequent Itemsets. *R package version 1.0-12*.
- Hayes, M. J. 2000. *Revisiting the SPI: Clarifying the Process*. Lincoln, NB: Drought Network News.
- Helsel, D. R., and R. M. Hirsch. 2002. Statistical Methods in Water Resources of Water Resources. In *USGS Techniques of Water-Resources Investigations Reports, Book 4, Chapter A3*. Vol. 4. Washington, DC: US Geological Survey.
- Herrmann, S. M., and K. I. Mohr. 2011. A Continental-Scale Classification of Rainfall Seasonality Regimes in Africa Based on Gridded Precipitation and Land Surface Temperature Products. *Journal of Applied Meteorology and Climatology* 50:2505–2513.
- Hillbruner, C., and G. Moloney. 2012. When Early Warning Is Not Enough—Lessons learned from the 2011 Somalia Famine. *Global Food Security* 1(1):20–28.
- Hipp, J., U. Guntzer, and G. Nakhaeizadeh. 2000. Algorithms for association rule mining: A general survey and comparison. *SIGKDD Explorations Newsletter* 2(1):58–64.
- Hornsby, C. 2013. *Kenya: A History Since Independence*. London, UK: I.B. Tauris & Co. Ltd..
- Kermanshahi. 2010. Gains made by al-Shabaab since January 2009. *Wikipedia*. http://en.wikipedia.org/wiki/File:Shababgains_2010.png. (accessed 15 October 2013).

- Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar. 2000. A catchment-based approach to modeling land surface processes in a general circulation model. *Journal of Geophysical Research* 105(D20):24809–24822.
- Kovac, N. 2013. *Rebuilding Somalia: Security Challenges for the Post-Conflict Nation*. Norfolk, VA: Civil-Military Fusion Centre.
- Little, P. D. 2008. Livelihoods, assets and food security in a protracted political crisis: The case of the Jubba Region, southern Somalia. In *Beyond Relief: Food Security in Protracted Crises*, ed. L. Alinovi, G. Hemrich, and L. Russo. Warwickshire, UK: Practical Action Publishing.
- Lu, X., L. Bengtsson, and P. Holme. 2012. Predictability of population displacement after the 2010 Haiti earthquake. In *Proceedings of the National Academy of Sciences* 109(29):11576–11581.
- Majid, N., and S.n McDowell. 2012. Hidden dimensions of the Somalia famine. *Global Food Security* 1(1):36–42.
- Maxwell, D., and M. Fitzpatrick. 2012. The 2011 Somalia famine: Context, causes, and complications. *Global Food Security* 1(1):5–12.
- Maxwell, D., N. Haan, K. Gelsdor, and D. Dawe. 2012. The 2011–2012 Famine in Somalia. *Global Food Security* 1(1):1–4.
- McKee, T. B., N. J. Doesken, and J. Kleist. 1993. The Relationship of Drought Frequency and Duration to Time Scales. In *Eighth Conference on Applied Climatology, 17–22 January, Anaheim, CA*, 179–184. Boston, MA: American Meteorological Society.
- McLeod, A. I. 2011. Kendall: Kendall rank correlation and Mann-Kendall trend test. *R package version 2.2*.
- Meleagrou-Hitchens, A., and H. Solomon. 2012. Factors responsible for al-Shabab's losses in Somalia. *CTC Sentinel*, 26 September. Combating Terrorism Center at West Point. <http://www.ctc.usma.edu/posts/factors-responsible-for-al-shababs-losses-in-somalia> (accessed 17 October 2013).
- Minnesota Population Center. 2011. *Integrated Public Use Microdata Series, International: Version 6.1*. Minneapolis, MN: University of Minnesota.
- Mishra, A. K., and V. P. Singh. 2010. A review of drought concepts. *Journal of Hydrology* 391(1–2):202–216.
- National Research Council. 2013. *Climate and Social Stress: Implications of security analysis*. Washington, DC: National Academies Press.
- Office of the Director of National Intelligence. 2012. *Global Water Security*. Washington, DC: Office of the Director of National Intelligence.
- Opiyo, D. 2011. Kenya Drought: Government Concedes Slow Response. *Daily Nation*. <http://www.nation.co.ke/News/-/1056/1108594/-/111rrs1z/-/index.html> (accessed 1 June 2013).

- ORNL. 2013. *LandScan Documentation*.
http://web.ornl.gov/sci/landscan/landscan_documentation.shtml#01 (accessed 20 December 2013).
- R Core Team. 2012a. *Auto- and Cross- Covariance and -Correlation Function Estimation*. <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/acf.html> (accessed 12 March 2013).
- R Core Team. 2012b. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, A. Touré. 2011. Assessment and Enhancement of MERRA Land Surface Hydrology Estimates. *Journal of Climatology* 24(24):6322–6338.
- Rienecker, M.M., M. J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M. G. Bosilovich, S. D. Schubert, L. Takacs, G. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, W. Gu, J. Joiner, R. D. Koster, R. Lucchesi, A. Molod, T. Owens, S. Pawson, P. Pegion, C. R. Redder, R. Reichle, F. R. Robertson, A. G. Ruddick, M. Sienkiewicz, and J. Woollen. 2011. MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climatology* 24(14):3624–3648.
- Robinson, C., L. Zimmerman, and F. Checchi. 2014. *Internal and External Displacement among Populations of Southern and Central Somalia Affected by Severe Food Insecurity and Famine during 2010-2012*. Washington, DC: Famine Early Warning Systems Network. http://www.fews.net/sites/default/files/documents/reports/Somali%20Displacement%20Report_FINAL_9Feb2014.pdf.
- Ross, W. 2011. Was the food crisis in Kenya avoidable? *BBC News*, 9 August. <http://www.bbc.co.uk/news/world-africa-14458109> (accessed 9 May 2013).
- Sailer, R. A., M. Hayes, and L. Bressan. 2002. Using the standardized precipitation index for flood risk monitoring. *International Journal of Climatology* 22(11): 1365–1376.
- Tatalovic, M. 2011. Forecasters 'Warned of Horn of Africa Drought' Last Year. *SciDev.Net* 14 July. <http://www.scidev.net/en/climate-change-and-energy/news/forecasters-warned-of-horn-of-africa-drought-last-year.html> (accessed 9 May 2013).
- UNHCR. 2012. *UNHCR Global Report 2010: Somalia*. Geneva: United Nations High Commissioner for Refugees.
- UNHCR. 2013. *Somalia Population Movement Trends*. United Nations High Commissioner for Refugees.
- UN-OCHA. 2011. *Horn of Africa Crisis: Situation Report No. 13*. New York, NY: UN Office for Coordination of Humanitarian Affairs.
- UT-Battelle LLC. 2010. *LandScan 2010 High Resolution Global Population Data Set*. Oak Ridge, TN: Oak Ridge National Laboratory.

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 04-21-2014		2. REPORT TYPE Technical Report/Final		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Socio-economic Effects of Drought in the Horn of Africa				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Jeanne M. Roningen and John B. Eylander				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Cold Regions Research and Engineering Laboratory (CRREL) US Army Engineer Research and Development Center 72 Lyme Road Hanover, NH 03755-1290				8. PERFORMING ORGANIZATION REPORT NUMBER ERDC/CRREL TR-14-5	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) The National Geospatial-Intelligence Agency 7500 GEOINT Drive Springfield, Virginia 22150-7500				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT					
13. SUPPLEMENTARY NOTES National Geospatial-Intelligence Agency Military Interdepartmental Purchase Request No: NIB8G12206GSA9					
14. ABSTRACT In recent years, the intelligence and defense communities have indicated interest in understanding the potential relationships between anomalous climate events and socio-economic consequences outside of the United States that could have implications for US national security. Our research evaluates potential linkages between retrospective climate analyses and empirical socio-economic datasets in Somalia and Kenya surrounding the 2011 drought in the Horn of Africa. Subnational-level data on internally displaced persons in Somalia from 2008–2012 were used to correlate drought-related population movements to climate-model-derived moisture indices. The analysis was expanded to account for livelihood zones and to investigate the predictive capabilities of linear models for observed population movements. Additional analyses investigated market price response to drought and market connectivity and explored the use of census data on household water infrastructure to assess drought vulnerability of specific communities. Results suggest that drought-induced migration response occurred with low but significant correlations across a broad range of medium- to long-term (6 months to 4 years) standardized drought indices but was limited largely to the geographic area in southern Somalia subject to a confluence of three factors: al-Shabaab governance during the 2011 drought, associated legal and operational impediments to aid delivery, and non-arid seasonality patterns.					
15. SUBJECT TERMS Climate Drought		Horn of Africa Internally displaced persons Market Prices		Migration Somalia Standardized Precipitation Index	
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT None	18. NUMBER OF PAGES 120	19a. NAME OF RESPONSIBLE PERSON
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)